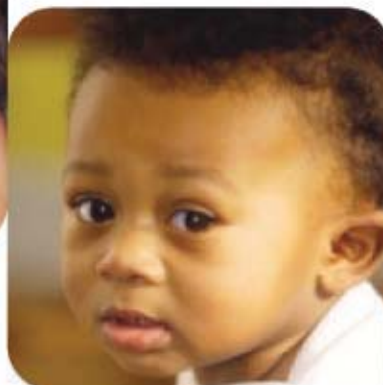


Preliminary Working Paper



RESOURCES FOR MEASURING SERVICES AND OUTCOMES IN HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

Prepared by:

Ellen Eliason Kisker
Kimberly Boller
Charles Nagatoshi
Christine Sciarrino
Vinita Jethwani
Teresa Zavitsky
Melissa Ford
John M. Love

Mathematica Policy Research, Inc.

Prepared for:

Rachel Chazan Cohen

Child Outcomes Research and Evaluation
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Abstract

This document contains resources to help Head Start programs that serve pregnant women and families with infants and toddlers develop a performance measurement plan and carry out data collection that will support their continuous program improvement efforts. These performance measures activities should build upon existing screening and assessment activities required by the Head Start Program Performance Standards. This document discusses the importance and development of a comprehensive plan and presents profiles of instruments that may be useful to programs. Because we anticipate that it will be used under different circumstances for different purposes, we present the background information in a book format, and the entire document on a compact disk, to enable users to search for the sections and measures that apply to them. This format will also support the continued evolution of the document, which is intended to be a “living” document to which information on new instruments can be added, through which new tools and approaches can be shared, and in which other resources that individual programs find useful can be compiled. These materials will be most useful when used in consultation with an assessment expert.

THE INCLUSION OF AN INSTRUMENT IN THIS RESOURCE DOCUMENT DOES NOT CONSTITUTE ENDORSEMENT OF THE INSTRUMENT BY THE AUTHORS, MATHEMATICA POLICY RESEARCH, OR THE U.S. GOVERNMENT.

Table of Contents

I. PERFORMANCE MEASUREMENT	1
II. FORMULATING A PERFORMANCE MEASUREMENT PLAN	7
III. INFORMATION INCLUDED FOR EACH INSTRUMENT	21
REFERENCES	33
APPENDIX A:	35
WORKSHEET TO HELP DEVELOP A COMPREHENSIVE PLAN FOR GATHERING AND ANALYZING DATA	
APPENDIX B:	39
PERFORMANCE MEASURES FRAMEWORK FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS	
APPENDIX C:	47
LIST OF MEASURES USED IN THE NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT	



Performance Measurement

Section 1



Performance measurement includes data collection and aggregation activities that give staff members the opportunity to look at how their program is doing, that is, to determine whether they are providing the services they intend to provide and how children and families are faring. Program staff can use this information in planning for continuous program improvement activities, and it can be shared with stakeholders such as parents and funders.

Head Start, as a national laboratory for early childhood education, has been a leader in developing outcomes-oriented accountability. It began developing performance measures in 1995, partly to respond to the Government Performance and Results Act (GPRA) requirements. In 1997, the Head Start Bureau launched the Family and Child Experiences Survey (FACES) to collect data on child and family outcomes, as well as program services and management systems, for a large nationally representative sample of children and families in Head Start programs (Administration

on Children, Youth and Families 2001b). This study links the development of children and families with their experiences in Head Start. Following the reauthorization of Head Start in 1998, Head Start programs were required to include child outcomes in their self-assessment process by 2003.

Although Early Head Start programs have not been required to report child outcomes, many have started to try to define and measure outcomes, for several reasons. Some programs are doing so in conjunction with performance measurement in Head Start because they operate within Head Start programs. Some are responding to other funders' requirements. Finally, some simply want to improve their services to families with infants and toddlers. In Spring 2001, the Early Head Start Technical Work Group recommended that the Head Start Bureau move forward to develop performance measures for Head Start programs serving infants and toddlers (Early Head Start and Migrant Head Start programs) to support programs' efforts.

The development of new performance measures for Head Start programs serving pregnant women and families with infants and toddlers has two purposes. These are (1) to create performance measures sensitive to the developmental stages from pregnancy to age 3 that can be used at both national and local levels for learning about

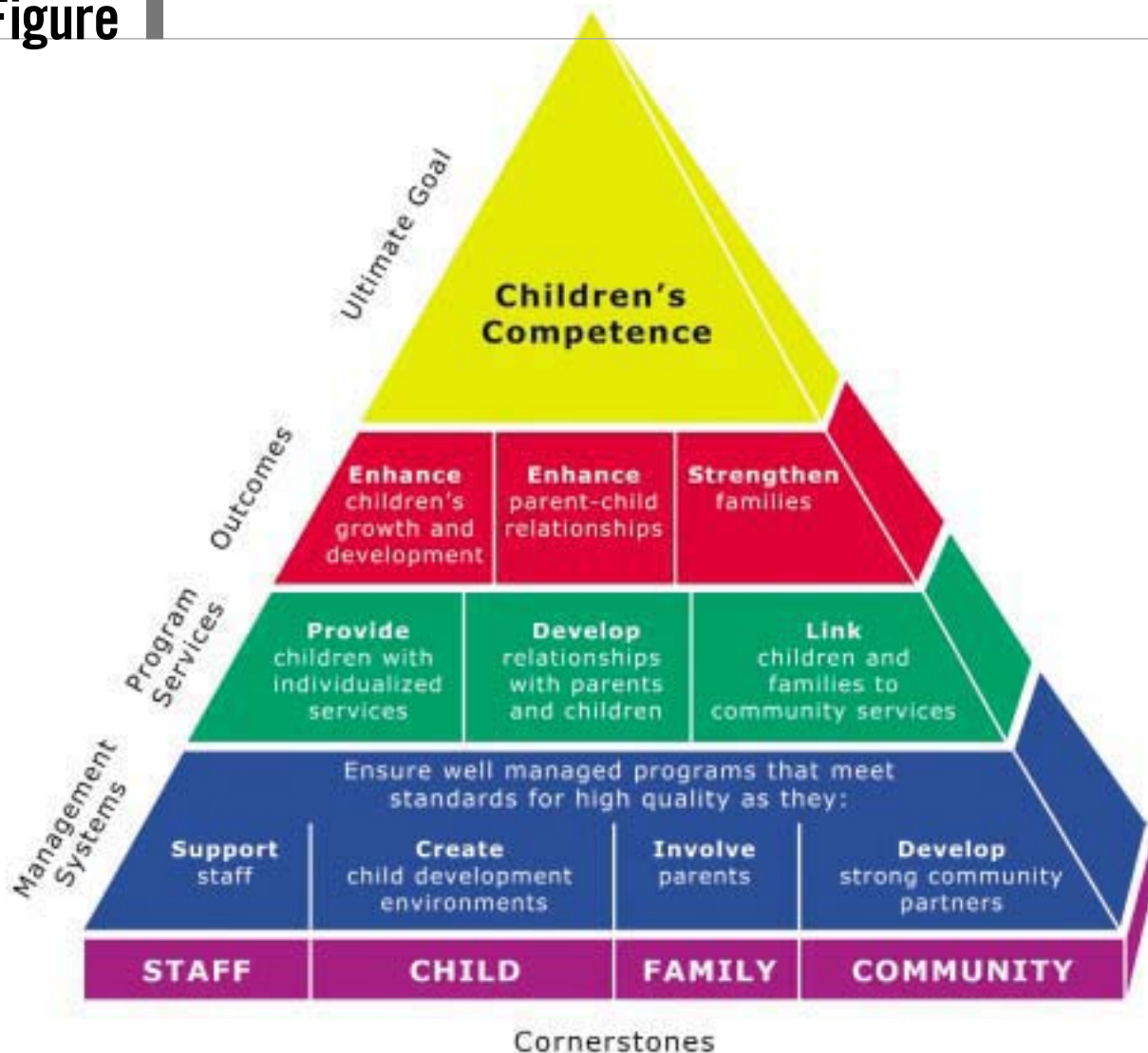
child and family experiences along with program services and management systems; and (2) to provide guidance to local programs in defining and measuring outcomes and using this information for continuous program improvement.

To fulfill these purposes, the Head Start pyramid, representing the Head Start performance measures framework, was expanded to reflect the unique features of programs serving infants and toddlers (see Figure 1 and Appendix C). The ultimate goal of Head Start programs that serve pregnant women and families with infants and toddlers is the same as that of Head Start programs that serve preschool children—enhancing children's competence.¹ The main outcomes supporting this ultimate goal—the blocks in the pyramid—have been expanded to reflect the central role of relationships in supporting attainment of this goal in families with younger children (U.S. Department of Health and Human Services [DHHS] 1994). Enhancing children's growth and development, enhancing parent-child relationships, and strengthening families as primary nurturers of their children are outcomes that support the ultimate goal of enhancing children's competence. Supporting these outcomes are objectives related to program services, including the objectives of providing children with individualized services, developing relationships with parents and children, and linking children and families to

¹ The Head Start goal, traditionally stated as children's "social competence," was shortened to "competence" because the interpretation of social competence in the Head Start pyramid has sometimes been narrower than was intended. Competence is the child's everyday effectiveness in dealing with his or her present environment and later responsibilities in school and life.

Figure 1

Conceptual Framework for Programs Serving Infants and Toddlers



needed community services. The objective of ensuring well-managed programs that meet standards for high quality in supporting staff, providing early childhood environments, involving parents, and developing community partnerships serves as the foundation of the pyramid. The pyramid rests on the four cornerstones recommended by the Advisory Committee on Services for Families with Infants and Toddlers—staff development, child development, family development, and community building (DHHS 1994).

The national Early Head Start Research and Evaluation project studied early program implementation and impacts on children and families (see Box 1). From this study, we learned that each program operates with its own theory of change—that is, a theory explaining how the services they provide will improve the child and family outcomes they are focusing on to meet the needs of children and families in their community. Although all programs must implement all aspects of the pyramid, the ways that programs



configure their management systems and program services (the bottom levels of the pyramid) vary widely. The specific child and family outcomes they focus on (the upper levels of the pyramid) also vary. As the program staff learns from continuous program improvement activities or as families' needs change, programs' theories of change may evolve. The evaluation showed that in general, patterns of program impacts reflected differences in theories of change. Home-based programs, which emphasized improving parenting and the home environment as an important path to improved child outcomes, had significant favorable impacts on parenting and child outcomes. Center-based programs, which emphasized direct services to children to improve outcomes, had favorable impacts on children and fewer significant impacts on parenting. The study also showed that programs that fully implemented key elements of the Head Start Program Performance Standards (elements related to all levels of the pyramid) had more favorable impacts on a wide range of outcomes than programs that were incompletely implemented (Administration for Children and Families 2002).

A *theory of change* is a belief or set of beliefs about how program services and other factors produce changes in the desired outcomes. The theory of change is the basis for designing and implementing program services. It may be explicit in program documents or staff discussions, or it may be implicit in the decisions of program designers and implementers.

The pyramid provides a framework to help programs identify their own theory of change. Together, the pyramid and theory of change provide guidance for examining links between program services and outcomes to inform continuous program improvement activities. The performance measures framework is general and does not identify specific program services and outcomes that should be measured. Individual programs must identify specific services and outcomes based on their own theory of change and select appropriate instruments to fit their available resources.

We do not recommend that programs collect data related to every element of the pyramid (at least to start with). We recommend that programs collect data related to each level of the pyramid (management systems, services, and outcomes), because having such information allows staff to link information on program characteristics and services with outcomes for children and families and to learn about both how well children and families are doing and how services might be improved to promote better outcomes.

Selecting appropriate measures is an important and complex process. This document is intended to be a resource for programs undertaking this process. It includes both screening instruments and assessment instruments that may be useful to programs as they explore how to approach performance measurement. We attempted to identify a wide range of instruments and select those likely to be most useful to Head Start programs serving pregnant women and families with infants and toddlers, but there are useful

tools that are not included in this review. For example, some kinds of assessment that are useful for individualizing services, such as portfolio assessment, are not included (Martin 1999). In addition, new measures are constantly emerging, and very recent measures may not be found here. Some programs may wish to use this resource document in consultation with an expert on screening and assessment to help them develop a plan and select instruments. (For more explanation of how we selected instruments for this resource document, see page 17.)

This resource document will be most useful to programs if it is used in conjunction with a comprehensive plan for performance measurement. In the next section, we discuss the importance of developing a plan and the elements that make up such a plan. In the final section, we describe how instruments were selected for this resource document, explain the information provided for each instrument, and present profiles of instruments that may be useful to Head Start programs that serve pregnant women and families with infants and toddlers.

Box 1

MAKING A DIFFERENCE IN THE LIVES OF INFANTS AND TODDLERS AND THEIR FAMILIES: THE IMPACTS OF EARLY HEAD START

A rigorous evaluation of Early Head Start services in 17 programs selected from the first groups of programs funded showed they had significant favorable impacts on a wide range of parent and child outcomes, some with implications for children's later school success. Findings from the study (*Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start*), using data gathered when children were age 3 and had completed the program, show that the programs sustained and broadened the pattern of impacts reported when children were age 2 (*Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families*, 2001). All Early Head Start evaluation reports are available online at [http://www.acf.hhs.gov/programs/core/ongoing_research/ehs/ehs_intro.html].

Early Head Start Improved Outcomes. The national evaluation conducted by Mathematica Policy Research, Inc. and Columbia University's Center for Children and Families at Teachers College, in collaboration with the Early Head Start Research Consortium, reported that 3-year-old Early Head Start children performed significantly better on measures of cognitive, language, and social-emotional development than a randomly assigned control group. While children who participated in Early Head Start performed better than their peers who did not receive Early Head Start on all aspects of development that were assessed, both groups lagged behind 2- and 3-year-old children nationally. This may, in part, be due to the fact that Early Head Start programs must reserve at least 10 percent of their slots for children with disabilities, including those with developmental delays who score at the lower end of the distribution. Parents in Early Head Start scored significantly better than control group parents on measures of many aspects of the home environment and parenting behav-

ior. Furthermore, Early Head Start programs enhanced parents' progress toward self-sufficiency years. Early Head Start fathers benefited as well. Although the overall impacts of Early Head Start were generally modest, the pattern of favorable findings across outcomes in a wide range of key domains is promising.

Full Implementation Was Important. The impacts on children and parents were consistent with the substantial difference the programs made in families' receipt of services. Early Head Start families were significantly more likely than control families to receive a wide variety of services, but especially to receive intensive services, and to receive intensive services that focused on child development and parenting.

Implementing the Head Start Performance Standards early and well is important for maximizing impacts on children and families. The research programs were systematically rated according to the extent to which they implemented key elements of the Performance Standards. Those that fully implemented the standards demonstrated a broader pattern of significant impacts than did the programs that did not reach full implementation during the evaluation period.

Patterns of Impacts Were Consistent with Theories of Change. Analyses that compared the contribution of impacts on parenting when children were age 2 to impacts on child outcomes at age 3 in programs providing home-based or a mix of home- and center-based services generally provided support for the theories of change that staff in those programs described, with some of the effects of the program on 3-year-old children being associated with the effects on parenting when they were age 2.



FORMULATING A PERFORMANCE MEASUREMENT PLAN

Section 2



The Head Start Program Performance Standards require programs to conduct screening and assessment activities and emphasize their importance for individualizing services and informing continuous program improvement. The performance standards allow considerable flexibility in how programs meet the requirements. In response to the requirements in the performance standards, Head Start programs serving pregnant women and families with infants and toddlers are already collecting data. Head Start programs serving pregnant women and families with infants and toddlers may use screening and assessment instruments for different purposes, including:

- **To Support Development and Learning.** Child assessment results can tell caregivers and teachers what each child can do and what he or she is ready to learn next. Family or parent assessment results can help program staff identify family or parent strengths, needs, and concerns and tailor services to the family. Over time, assessment results can demonstrate how each child and family is progressing. Assessment results can also help staff communicate with family members

about their children's needs and progress, as well as their own needs and progress. This information can help staff individualize services and improve them over time.

- ***To Identify Special Needs and Concerns.***

Because of the cost of in-depth assessments, screening is usually the first step in identifying special needs. Children or families for whom an in-depth assessment is indicated are often referred to a physician or other expert for a complete evaluation. Screening results provide the information needed for referrals to other agencies, such as Part C agencies, to obtain services for children with disabilities.

- ***To Evaluate the Program and Monitor Trends.***

For this purpose, child and family screening and assessment data may be aggregated and used to inform continuous program improvement efforts. Program assessment data and feedback from Head Start Bureau monitoring may also be used for this purpose. Aggregated screening and assessment results can inform staff about how well the program is meeting child, family, and community needs.

Screening and assessment results for individual children, along with other information from parents and caregivers, are needed to help staff tailor services for those children. When the same instruments are used for all children in a program, aggregating data across families can provide a picture of how children and families in the program are doing overall. When this aggregate information on child and family outcomes is linked to information on services and other pro-

gram characteristics, it can provide insights that are useful to staff members in their continuous program improvement efforts. Aggregating data provides a picture of how children and families and the program are doing (that is, it measures the program's performance). Over time, aggregated data can be used to track changes in child and family functioning, which along with information about changes in program services or characteristics, can help staff learn about what works best for particular types of children and families. The aggregated information on how children and families in the program are doing can also be useful for meeting other funders' reporting requirements.

Screening. Screening is a generic term referring to activities designed to identify individuals who have a high probability of exhibiting delayed, abnormal, or problematic development. The screening is intended to identify problems at an early stage and identify individuals for whom further, in-depth assessment activities are needed.

Assessment. Assessment is a generic term referring to procedures for obtaining systematic information on a child's, parent's, family's, or program's strengths or needs. As noted in Chapter I, the Head Start Program Performance Standards focus on the child and family assessment purposes of identifying "(i) the child's unique strengths and needs and the services appropriate to meet those needs; and (ii) the resources, priorities, and concerns of the family and the supports and services necessary to enhance the family's capacity to meet the developmental needs of their child."

[See the Early Head Start National Resource Center's Technical Assistance Paper #4 for more detail on screening and assessment activities.]

Performance measurement includes data collection and aggregation activities that give staff members the opportunity to look at how their program is doing, that is, to determine whether they are providing the services they intend to provide and to learn how children and families are faring. Program staff can use this information in planning for continuous program improvement activities, and it can be shared with stakeholders such as parents and funders.

Screening results for individual children and families are useful for deciding whether further assessment is needed. Screening results can also be aggregated to provide information on the extent of potential problems in the population and the need for in-depth assessments among children and families overall. Aggregated screening results and information on referrals can inform a program as to whether these in-depth assessments are happening if the program's tracking system does not provide this information.

Assessment results for individual children can be used for planning services; in addition, they can often be aggregated to provide broader information on child outcomes. Some approaches to assessment are valuable for individualizing services, but cannot be aggregated unless they are translated into another form. For example, portfolio assessment can be extremely useful for individualizing services for children, but unless a systematic

way of coding the information is developed, the results cannot be aggregated.

Programs face both a great opportunity and a significant challenge as they consider performance measurement. They have an opportunity to select instruments and collect data that best meet the needs of their families and their program, and they face the challenge of figuring out just what those instruments should be, how they should be administered, and how the information collected using those instruments should be analyzed. Meeting that challenge effectively requires programs to develop a plan that considers the purpose(s) for which the data will be used, what data are already being collected and additional data that need to be collected, and how data will be aggregated and analyzed for continuous program improvement. Taking the time to develop a comprehensive plan will help ensure that the program's resources for the required ongoing screening and assessment of individual children and families, as well as the new performance measurement and data aggregation activities, are used well and that the activities are as useful as possible.

To formulate a comprehensive plan, programs need a clear understanding of why performance measurement is needed. What does the staff need to know to determine whether the program is achieving its intended outcomes and to individualize or optimize services for children and families? What information is essential? What, in addition, would be good to know? Some questions programs might consider to help reach this understanding are:



- What are the program's intended outcomes? What is the program trying to accomplish?
- How will the program's services influence these outcomes? What will the program do or provide for its families to achieve the desired outcomes?
- What information does the program need for individual service planning? Is this information needed at enrollment? How often and at what times during the program is this information needed? What information is already being collected (such as HSFIS data, child and family screening/assessments) that can be used for this purpose?
- What information does the program need for program planning? How often is this information needed? What information is already being collected (such as HSFIS data, PIR data, program self-assessment information) that can be used for this purpose?
- How will the program know that staff are doing what they think they are doing?
- How will the program know how well children and families are progressing?
- Overall, are services implemented well? How will the program know services are being implemented well?
- Is the program influencing the targeted outcomes across all families served? Are particular types of families benefiting more (or less)?

- What information will help demonstrate to funders that the program is effective?

- How will the program use the information gathered for program improvement?

To make performance measurement activities as useful as possible, programs need a theory of change. This model, or theory of change, provides a framework for guiding the selection of individual instruments and for integrating information obtained from the selected instruments and other sources of information about a child and family. Such a model or theory of change simply specifies explicitly what child and family development outcomes the program is trying to improve and how the program's services will influence those outcomes (comprehensive answers to the first two questions in the list above constitute a theory of change).² With such a model or theory of change, the staff can select instruments that will focus on measuring targeted outcomes and the key services designed to improve them. The program's theory of change can also guide the interpretation of data at both the individual and program levels and decisions about what to do in response to the information.

The performance measures pyramid, along with a theory of change, provides a framework for developing a program-specific comprehensive plan for performance measurement. We do not recommend that programs collect data related to

²For an illustration of the application of the theory-of-change approach to identifying expected outcomes in Early Head Start, see the first report on program implementation of the national Early Head Start Research and Evaluation project (ACYF 1999, Chapter II).

every element of the pyramid (at least to start with). We recommend that programs collect some information related to each level of the pyramid (management systems, services, outcomes, and the ultimate outcome), because having such information allows staff to link information on program characteristics and services with outcomes for children and families and learn about both how well children and families are doing and how services might be improved to promote better outcomes. All plans should include measurement of child outcomes.

The following example illustrates how the performance measures framework and an explicit theory of change can help program staff develop a comprehensive plan for gathering and analyzing information. A program that provides home-based services may have a theory of change that indicates that it is trying to improve child development outcomes indirectly by improving parenting and parent-child relationships. To accomplish this, it provides home visits in which staff members work with parents on specific parenting skills. The program may emphasize improving children's language development and train home visitors to work with parents on activities they can do with their child to promote learning and language development, such as reading to them regularly and providing a home environment that supports learning and language development. Based on its theory of change and emphasis on language development, this program might give priority to selecting instruments that measure children's language development (outcome), the frequency of

parents' reading to their child (outcome), support for language and learning in the home environment (outcome), the frequency and quality of home visitors' interactions with families focused on children's language development (services), and supervisory practices to support home visitors' interactions with families (management system). Ideally, the data obtained with these instruments will indicate how children's language development is progressing, provide insights into how parents' progress in parenting skills may be promoting their children's language development, illuminate how program services may be contributing to parents' progress (and indirectly to children's language development), and indicate how home visitor supervision is supporting home visitors' activities related to language development. The data may also provide insights into ways that home visitors' activities with parents can be improved to enhance children's language development further.

A comprehensive plan for gathering and using data includes several important elements.

The following questions indicate important elements that a comprehensive plan should include:

- Does the plan include clearly stated purposes for gathering and using data?
- What instruments will be used to gather information for individual service planning and aggregate performance measurement?
- With which children or families will each instrument be used?
- When and how often will each instrument be administered?

- Who will administer each instrument, and what training will they receive?
- How will the results of administering each instrument be recorded?
- How will the results of each instrument be kept confidential?
- How will the results be shared with parents?
- How will results be used to plan services for individual children and families? Will they be used for referrals to other community service providers, such as Part C agencies?
- Will the individual results be aggregated across children and families and analyzed for program planning? If so, how?
- How will the results feed into continuous program improvement?
- How will the results be reported to other stakeholders?
- How will the plan be modified as you gain experience with it?

Appendix A contains a worksheet that provides one way to summarize some of this information and that may be helpful as a component of a comprehensive plan.

In developing a plan that is appropriate and feasible for its program, the staff needs to consider its priorities among information needs.

Programs may not have the resources to gather and analyze all useful data, and staff members may not have experience with aggregating results from screening and assessment instruments. Although it is important for programs to create a comprehensive plan, it may be necessary, for guiding program activities, to identify priorities that are feasible with current resources and staff experience.

As staff gain more experience with collecting and aggregating data, additional elements of the plan can be implemented (if resources are available).

The kinds of resources that should be considered in developing and implementing a comprehensive plan for gathering and analyzing information include staff resources, technical resources, and financial resources. Staff resources include the time that could be devoted to administering instruments and analyzing the results, as well as activities to build skills and obtain training. Technical resources include training and technical assistance, consultation, and computer hardware and software—as well as qualified personnel—to manage and analyze data. Financial resources include money to purchase needed materials, pay for additional training and technical assistance, and purchase computer hardware or software.

In determining measurement priorities, programs should consider giving priority to instruments that together represent a balance across elements of the pyramid for Head Start programs serving pregnant women and families with infants and toddlers. As noted earlier, lower entries in the pyramid (program management and services) support the outcomes above them. Thus, information on program management and services along with information on child and family outcomes will provide greater insights into ways in which program services can be improved to promote better outcomes at both the individual and the program levels. First priority, however, should be given to measuring child outcomes.



A comprehensive plan for gathering and analyzing data will be useful only if the staff is willing and able to implement it. In developing such a plan, program managers need to build support among staff members and prepare them to use the selected instruments. Involving key staff members who will implement the plan in its development may promote their “buy-in” to the new activities. Involving staff will also help to illuminate the questions that the program most needs to answer. In addition, providing enough training and time to administer the selected instruments may also help staff members embrace the new activities. Finally, involving staff members in interpreting and using the results will help them see how they are useful, motivate them to administer the instruments well, and enable them to help identify needed modifications to the plan.

A key part of the development of a comprehensive plan is the selection of specific instruments that will be used. The checklist in Box 2 lists some questions to help the program staff identify instruments that will best meet its needs. In addition to these questions for considering individual instruments, it is important that, together, the selected instruments draw on multiple sources of information. To be useful for aggregating across families, they also need to be administered to all children or families (or the group of children or families for whom aggregated information is needed) in a consistent manner.

In developing a comprehensive plan, care should be taken to prevent misuses of instru-

ments. It is not appropriate to select a subset of items from an instrument, combine items from multiple instruments, or change the wording or response categories for items in an instrument, because the abbreviated or changed instrument may not be reliable or valid. Some instruments, however, include official subscales or subtests that may be used alone. Only if the directions for using an instrument indicate that using just a subscale or subtest is appropriate should staff select and use parts of instruments. To prevent misuse, it is also important that staff members who will be administering an instrument and interpreting the results have sufficient knowledge and training to enable them to do so accurately and appropriately. Lack of knowledge and understanding of an instrument can lead to its misuse.

While implementing a plan for gathering and analyzing data, a program may see the need for changes to the plan. The instruments selected initially may not work well (for example, requirements for administering them may be too difficult to meet or scoring may be too difficult), and different instruments may meet program needs better. The staff may also find that instruments selected initially do not provide all the information needed and that alternative or additional instruments may better meet the program’s needs.

The development and implementation of a plan for gathering and analyzing data takes time. Figure 2 illustrates the development of a continuous improvement model in the Clayton Family Futures program in Denver, Colorado. It

summarizes the steps that the program has taken to develop its model, the resources required, the timeframe for each step, and the implications of each step for the program. Over time, the program's continuous improvement activities have grown as the staff has experienced the value of the information and asked more questions about program services and how children and families are doing.

A plan for measuring outcomes can also be implemented at a broader level. Box 3 illustrates a statewide plan for collecting and analyzing data on a common set of performance measures. State Early Head Start Programs in Kansas have agreed to collect data on a common set of measures to support outcomes-based community planning.

Box 2

CHECKLIST TO HELP IDENTIFY APPROPRIATE SCREENING AND ASSESSMENT INSTRUMENTS

Below are elements of screening and assessment instruments and their use that contribute to their usefulness for Head Start programs serving pregnant women and families with infants and toddlers. For each instrument under consideration, check the box beside each element that applies. The more boxes that are checked, the better is the match between the program's needs and the instrument.

	Instruments					
The instrument measures what the program wants to know.						
The instrument was designed for the purpose for which it will be used.						
The instrument is appropriate for the cultural backgrounds of children/families who will be assessed.						
The reliability and validity of the instrument are sufficiently high for the purposes for which it will be used.						
Sufficient resources are available to obtain and use the instrument.						
It is feasible to administer the instrument according to the instrument developer's directions.						
The instrument facilitates sharing information about children (or families) with staff and parents.						
Staff members who will administer the instrument have (or will receive) the training needed to administer and score the instrument correctly.						
The instrument is appropriate for children with disabilities (or their parents).						
It is feasible to administer the instrument in settings children (or families) are comfortable with.						

Figure 2

THE DEVELOPMENT OF A CONTINUOUS IMPROVEMENT MODEL CLAYTON FAMILY FUTURES EARLY HEAD START PROGRAM

<p>Frequently Asked Questions</p> <p>Will what works in one program system work in other areas?</p> <p>Ex. The system for providing Diet and Nutrition screenings was identified as very efficient. Could we apply this system to other program areas needing a similar process?</p> <p>What actions can we take in areas that are not reaching the benchmarks established?</p> <p>Several internal actions were taken to address the challenges in providing dental screenings. However, they did not produce the desired outcomes. In response, the health team created a strong community collaboration with the School of Dentistry that was extremely effective in connecting children to dental services.</p>	<p>Step One: Setting up tracking systems and monthly reporting formats for a limited number of outcomes (a good starting point is to ask, “Are we meeting the Performance Standards in all areas?”)</p>			
	<p>Actions</p> <ul style="list-style-type: none"> Identify what you want information about Identify what is currently being tracked in those areas Identify the data tracking forms that exist and/or that need to be created Set up databases to support the collection of the information Set up monthly summary reporting formats and ongoing deadlines for the report distribution 	<p>Resources</p> <ul style="list-style-type: none"> Analyze the budget – begin to allocate funds for supporting continuous improvement implementation (start small) Designate existing and/or new staff to carry out the actions 	<p>Time Frame</p> <p>2-3 months</p>	<p>Program Implications</p> <ul style="list-style-type: none"> Questions about the link between reports and job performance Training for supervisors to effectively utilize the reporting in supervision Creation of meaningful dialogue about barriers encountered by staff in delivering quality services
<p>Developed by Chris Sciarrino, The Clayton Foundation, Denver, Colorado</p> <p>Consultation by Charmaine Lewis, Clayton Family Futures Early Head Start, Denver, Colorado</p>	<p>Step Two: Beginning to use inquiry and analysis as a method of self-evaluation, reflection, and program improvement</p>			
	<p>Actions</p> <ul style="list-style-type: none"> Disseminating summary reports to appropriate staff Program leadership leads the way for team analysis by working with the creator of the reports to identify trends, issues, and strengths. Monthly team analysis meetings are established with key staff responsible for supervision of program implementation areas. Action plans are created with time lines to address identified areas of need and how reports will be used in supervision Establishment of benchmarks for every outcome area 	<p>Resources</p> <ul style="list-style-type: none"> This model utilized 1 FTE for supporting the development and ongoing implementation of the design One formal meeting for the director to discuss the reports with the creator of the reports was established A monthly administrative team meeting focused on analyzing the reports was established Several informal discussions regarding implementation, analysis, and supervision strategies were needed to modify and adapt the approach 	<p>Time Frame</p> <p>1 year</p>	<p>Program Implications</p> <ul style="list-style-type: none"> Working with program coordinators to see reports as objective and representative of what families and children are receiving versus “what I have not done.” Beginning to use a reflective process to create change Ability to use tangible evidence is an opportunity to acknowledge and address needs Creating pride in accomplishment of benchmarks

“Having information about outcomes reported to me has really helped me plan for my whole class and individual children.” “Having information about outcomes reported to me has really helped me plan for my whole class and individual children.”

Figure 2

Frequently Asked Question	Step Three: Using Continuous Improvement data for program planning and communication with stakeholders			
	Actions	Resources	Time Frame	Program Implications
<p>Is it all about the numbers?</p> <p>Staff struggled with putting quality into numbers. Herein lies one of the basic challenges of the approach—how is it done so it is not seen as an either/or paradigm, either quality or accountability? It is a tendency to see these as opposing views. It is our belief that these are not separate concepts. Being able to provide consistent services (accountability) is basic to the quality and integrity of the program. It is the analysis of the data and how leadership is able to interpret its impact on quality that brings the process to life.</p>	<ul style="list-style-type: none"> • Utilize reports to draw conclusions and ask questions about the EHS experience for children and families. Link this information to reflective supervision with staff. • Reports collected over time are compared to demonstrate trends, highlight issues, program strengths and needs • Information is used in conjunction with yearly self-assessment and community needs assessment for program planning • Reports continue to be refined as adaptations needed present themselves. <ul style="list-style-type: none"> ■ Reports are summarized quarterly for dissemination to stakeholders 	<ul style="list-style-type: none"> • Meeting time • Staff commitment to a set of clearly articulated and shared values- <ul style="list-style-type: none"> ■ Quality-commitment to striving for excellence; doing the best possible job working toward a common vision ■ Accountability-commitment to a set of clear, well defined and high standards (i.e., performance standards); demonstrating through action our ability to meet those standards ■ Openness and collaboration-commitment to and examination of diverse perspectives and engaging in group processes and partnerships that help determine the path to high quality programs ■ Reflection- commitment to reviewing and dialoguing about current practices along with an acceptance of one’s personal responsibility in achieving high quality ■ Self-growth-dedication to each individual’s growth ■ Follow through-commitment to making happen what is planned, expected, and desired 	<p>ongoing</p>	<ul style="list-style-type: none"> • As coordinators become invested in and comfortable with the reports, they begin meeting together to ensure integration of Continuous Improvement efforts and to identify barriers to quality implementation of services. • Analyses discussions became more complex as staff began to ask more questions. • Having aggregated data readily available on a monthly basis allows for staff to easily answer 1.) Are we doing what we say we are doing? 2.) Are we accomplishing the outcomes we want to accomplish? • Creating a meaningful dialogue and sharing of outcomes information with governing boards, policy council, parents. Readily available outcome data allows for timely and accurate reports to funders
<p>Developed by Chris Sciarrino, The Clayton Foundation, Denver, Colorado</p> <p>Consultation by Charmaine Lewis, Clayton Family Futures Early Head Start, Denver, Colorado</p>				

Comments from staff:

“At first it felt vulnerable and scary that my job was out there in numbers, but now I realize how helpful it is to know just what is happening and how the reports help me to offer better services.”

“We know we are ‘walking our talk.’”

“Rather than responding to a checklist for someone else, we are using the information with staff and for planning.”

Box 3

AN EXAMPLE OF A STATEWIDE APPROACH TO MEASURING OUTCOMES FOR EARLY HEAD START

Efforts to measure and report outcomes can be implemented at broader levels. In Kansas, for example, state Early Head Start programs developed and agreed to collect data on a core set of outcomes as part of a statewide system for assessing services for children and families. This system, called Connect Kansas, supports outcome-based community planning and community capacity building to create and sustain environments in which all Kansas children are safe, connected, nurtured, and supported by caring and involved adults and communities.

Outcomes for Early Head Start programs were developed through four focus group discussions. These focus groups included a wide range of stakeholders, state administrators, federal Administration for Children and Families Region VII staff, Head Start Quality Improvement Center staff, Early Head Start directors, and parents.

The following core outcomes will be measured consistently by 13 state Early Head Start grantees in 32 counties. Families must be enrolled in Early Head Start for a minimum of 6 months to be included in outcomes measures.

Outcome 1: Pregnant women and newborns thrive.

___% of pregnant women who receive prenatal care within the first 45 days of enrollment

___% of new mothers who deliver an infant weighing 5.5 pounds or greater

Outcome 2: Infants and children thrive.

___% of teachers working on or having a minimum Child Development Associate (CDA) certificate Show Breakdown:

___ are working on a CDA (any stage but not yet credentialed)

___ have acquired a CDA

___ have an AA/AS in ECE or related field

___ have a BA/BS in ECE or related field

___ have a MA/MS in ECE or related field

___ have other degree, specify

___% of Early Head Start learning environments with a score of 5 or higher using the Thelma Harms Rating Scale (measured at entry, 6 months, 1 year, and every year thereafter. Data should be taken from the last score.)

___% of Early Head Start children who are up-to-date on immunizations

___% of Early Head Start children who are up-to-date on well child checks/Kan Be Healthy

Outcome 3: Children live in stable and supported families

___% of parents who demonstrate improved parenting skills (measured by the Parents as Teachers Parent Knowledge Questionnaire upon entry, 6 months, 1 year, and every year thereafter)

Box 3

___% of enrolled families with one or more parents employed, enrolled in school, or attending a job training program 9 out of 12 months enrolled in Early Head Start

Show Breakdown:

- ___ less than 30 hours of employment
- ___ greater than or equal to 30 hours of employment
- ___ enrolled in school (part time or full time)
- ___ attending a job training program

___% of families who have a supportive home environment for their child with a variety of learning experiences and materials (measured by the HOME upon entry, 6 months, 1 year, and every year thereafter. Data would be taken from the last score, middle half or higher. Data will not be collected on first time pregnant women until after the birth of the baby.)

Outcome 4: Children enter school ready to learn.

___% of children without a diagnosed disability who demonstrate age-appropriate development in the three domains of: Intellectual, Social-Emotional, and Motor Skills (measured by the Parents as Teachers Developmental Milestone Checklist)

___% of children who demonstrate age-appropriate language (as measured by the Early Communication Indicator, Juniper Gardens)

Other data needed for collection purposes only:

1. Total # of children enrolled in EHS who are receiving child care services.
2. Total # of non EHS children receiving quality child care services in EHS child care partnerships.
3. % of children identified through screening for further intervention services.
4. % of children with an IFSP, Individual Family Service Plan or IEP, Individual Education Plan, in place (IFSP/IEP denotes special services).

For additional information contact:

Mary Weathers

Kansas Early Head Start Manager
Social and Rehabilitation Services
915 SW Harrison, Room 681 W
Topeka, Kansas 66612
785-296-4712
mxkw@srskansas.org

Lynda Bitner

Administration for Children and Families
601 E. 12th Street, Room 276
Kansas City, Missouri 64106
816-426-5401
lbitner@acf.hhs.gov

INFORMATION INCLUDED FOR EACH INSTRUMENT

Section 3



The purpose of this resource document is to provide information, in one place, about existing screening and assessment instruments designed for use with children under age 3 and their families, as well as instruments designed for assessing services provided by programs serving them. Thus, we cast a broad net and include a wide range of screening and assessment tools of potential use to programs. Many of the instruments described are established instruments that yield a standard score that places the child's performance in the context of other children of the same age. We also include some data collection tools that may be useful, such as implementation rating scales and questionnaires that include questions on family practices, health, and health care receipt from the national Early Head Start Research and Evaluation Project.

We did not set strict inclusion criteria, but tried to provide information on a range of features for each instrument so programs can make informed decisions in selecting instruments. Each program must determine the purposes for considering a particular instrument and evaluate how well the instrument fulfills those purposes.

In general, because of their limited applicability for programs serving infants and toddlers, we did not include measures for which the lowest appropriate age for administration was older than 2 years. We made an exception for certain instruments, such as the Woodcock Johnson III and Peabody Picture Vocabulary Test, that Head Start programs sometimes use and that may be helpful for continuity when children go on to Head Start.

We consulted multiple sources of information to identify instruments for inclusion in this resource document. We looked at the National Early Head Start Research and Evaluation Project to identify instruments used by the national and local research teams and instruments that research programs used. We held group discussions with Early Head Start program staff at the 2002 Birth to Three Institute to learn about screening and assessment tools they are using in their programs. Information was provided about screening and assessment tools that Early Head Start programs are currently using. We consulted with researchers and technical assistance experts. Finally, we conducted a literature review to identify instruments that are used widely and have been developed and/or normed within the past 15 years, or after 1987.

The instruments included in this document were developed for a variety purposes and by individuals from different disciplines. Thus, you

may find that some instrument names are overly technical or offensive. In these cases, you may want to present the instruments to parents using a less technical name that describes what the instrument measures in terms that parents will understand. For example, you might want to refer to the Parent-Child Conflict Tactics Scale as a questionnaire on discipline and responses to children's behavior.

The screening and assessment instruments in this resource document are presented in three groups: (1) instruments for measuring child development; (2) instruments for measuring parenting, the home environment, and parent well-being; and (3) instruments for measuring program implementation and quality. Within each group, instruments are in alphabetical order. Summary tables listing the instruments are presented at the beginning of each group of instruments.³ This resource document is intended to be a living document that will be updated as new screening and assessment instruments are identified or become available.

We gathered information about each instrument from different sources, depending on the type of instrument. For the more formal, copyrighted instruments, we relied primarily on the manuals or Web-based information available from the authors or their publishers. If we found a key research article about a formal instrument, we

³ THE INCLUSION OF AN INSTRUMENT IN THIS RESOURCE DOCUMENT DOES NOT CONSTITUTE ENDORSEMENT OF THE INSTRUMENT BY THE AUTHORS, MATHEMATICA POLICY RESEARCH, OR THE U.S. GOVERNMENT.

also reviewed it and included the pertinent information. For the more experimental, less formal instruments, we reviewed the instrument itself and the supporting material we were able to locate, such as research reports and published articles, and reviews conducted by others. Each entry includes a reference section that identifies the sources of information we used.

Many of these instruments are grounded in developmental theory and research. Developers of standardized tests for children usually begin with their theory of how abilities develop and identify areas to be assessed. Then they create items to measure the identified areas and try them with children to determine whether the items discriminate among children by age. After a core set of items is identified, test developers often launch a large, nationally representative study to test the items and obtain statistical information about how the study participants performed on each item. From the study findings, the test developers determine the best set of items, develop rules about where to begin and end the test, and decide on procedures for converting raw scores (based on summing the number of items answered correctly or on the average rating across items on a rating scale) to norm-referenced scores. The norm-referenced scores take advantage of the nationally representative study and allow comparisons between how an individual child performed on the test and how children of the same age in the study performed. The nationally representative study also provides information about how the instrument works with diverse and low-income

populations.

Other types of research also provide important information about a screening or assessment instrument. Studies that use a new instrument in conjunction with established instruments that measure the same ability or skill provide information about whether the new instrument measures what it was intended to measure. Other studies compare how well the new instrument predicts children's performance in a given skill area many years later. Because they take a long time to conduct, these studies are not available for very new instruments, but they can be valuable in evaluating an instrument administered when children are young.

No screening or assessment instrument performs perfectly across all the dimensions practitioners and researchers believe are important (such as the statistical properties of the instrument or how easily the resulting information feeds back into individualized intervention planning) and for all the purposes for which the instrument may be used. We encourage you to weigh the information described for each instrument according to your program's theory of change, your comprehensive plan for gathering and analyzing data, and the purposes for which you will use the information. Consultation with an expert may help you sort through this information and select screening and assessment instruments.

The language that describes screening and assessment instruments is filled with jargon. Box 4 defines the key terms used in this document.



The rest of this chapter includes a summary of what you will find described for each instrument included in this resource document. Each entry includes a summary table and a more detailed description of the topics we identified as most useful for making comparisons across instruments. The topics in the summary table include:

- **Authors, Publisher, Ordering, and Initial Material Cost Information.** This information will allow you to obtain the instruments. Some publishers will provide an inspection copy of the materials for a short period of time at no charge. Some publishers require that only trained psychologists or other assessment professionals purchase and use the materials, because the content of the instruments must be kept confidential and the instruments must be administered and used in accordance with professional guidelines. We list the cost for the initial materials required to use the instruments. For some copyrighted materials, you will be required to purchase a score sheet for each screening or assessment you conduct. You may be able to negotiate with the publisher for a reduced price if you are buying in bulk.
- **Representativeness of Norming Sample.** As described in Box 4, knowing whether the norming sample was nationally representative or representative of the children or parents in your program is important in deciding whether to use an instrument. Your screening and assessment plan will include the purpose for each screening and assessment. If you are interested in how the chil-

dren in your program are performing compared with children nationally, you will want to choose an instrument with a nationally representative norming sample.

- Knowing how children from low-income families in the norming sample performed compared with all children nationally can also be important for interpreting assessment results. For example, the Early Head Start Research and Evaluation study found that children's standardized scores on the Bayley Mental Development Index decreased between 14 and 24 months of age and remained at the 24-month level at 36 months. This pattern has also been found in other studies of low-income children and in the Bayley norming sample. In this case, the decrease in standardized scores reflects differences in the composition of the test at different ages. At 14 months of age, the Bayley does not include many items directly focused on language development. At 24 and 36 months, the Bayley includes many items that tap language development. The decrease in standardized scores among low-income children as they get older indicates that low-income children score less well compared to children nationally as language development becomes a more important part of the test.
- **Languages.** We included the languages in which the instruments are available. Some instruments have unofficial translations used in the field, but we restricted our listing to the languages that are available from the authors or publishers. If you are planning to use an instrument to compare the children in your program with those in the

Box 4

BRIEF DEFINITIONS OF KEY TERMS

Assessment. Assessment is a generic term referring to a variety of procedures for obtaining systematic information on a child's, parent's, family's, or program's strengths or needs. As noted in Chapter I, the Head Start Program Performance Standards focus on the child and family assessment purposes of identifying "(i) the child's unique strengths and needs and the services appropriate to meet those needs; and (ii) the resources, priorities, and concerns of the family and the supports and services necessary to enhance the family's capacity to meet the developmental needs of their child." These two major purposes of assessment are sometimes described as providing information for individual diagnosis and program planning. The purposes of a diagnostic assessment are to (1) identify whether an individual has special needs, (2) determine what the problems are, (3) suggest the cause of the problems, and/or (4) propose strategies to address the problems (Meisels and Provence 1992). The purposes of an assessment for program planning are to (1) learn about an individual's ability to perform particular tasks or achieve mastery of particular skills, and (2) design intervention activities for the individual that support the completion of tasks and mastery of skills over time. Depending on the purpose of the assessment process, it may include norm-referenced tests; observations in the home, child care, early intervention, program, or school setting; interviews with family members, child care providers, or others who may provide important information about the individual; and ratings by adults knowledgeable about the child (including a parent, caregiver, or teacher) (Sattler 1992). The performance standards also require programs to conduct an "assessment of community strengths, needs, and resources," as well as an annual program self-assessment of "effectiveness and progress in meeting program goals and objectives and in implementing federal regulations."

Screening. Screening is made up of a set of activities designed to identify individuals who have a high probability of exhibiting delayed, abnormal, or problematic development. The screening is intended to identify problems at an early stage and to use this information to flag individuals for further, in-depth assessment activities.

Basal. A basal is established on a standardized test when the individual demonstrates that he or she successfully completes the first few items administered. On most standardized tests, the tester begins administering the items based on how old the individual is, starting later if the individual is older. If the individual passes the number of items specified in the test manual for establishing a basal, the tester is able to assume that the individual would have gotten all of the previous items correct and adds in the number of untested items to the correctly passed items administered to the individual. If the individual does not pass the specified number of items, the tester would administer earlier items until the prescribed number of items are passed or the tester reaches the start of the test. Using a basal rule saves time during the testing session and reduces fatigue.

Ceiling. A ceiling is established on a standardized test when the individual demonstrates that he or she fails a few of the later items administered. On most standardized tests, the tester continues administering the items until a certain number (either in a row or a proportion, such as six out of eight in a row) are failed. If the individual fails the number of items specified in the test manual for establishing a ceiling, the tester ends the test and is able to assume that all later test items would be failed by that individual as well. This saves time during the testing session and reduces fatigue.

Criterion-Referenced Test. This type of test compares an individual's performance to an established measure of performance rather than to the performance of others. Criterion-referenced tests will usually include a measure of mastery, or how well a child is able to complete a task. For example, if a test required that a child identify all of the letters of the alphabet, that would be a criterion-referenced test. We would be able to describe the child's mastery of the test by using statements such as, "The child is able to identify 80 percent of the letters in the alphabet."

Norm-Referenced Test. This type of test compares an individual's performance to the performance of others on the same measure. Usually, the norms are developed from data collected from a large, nationally representative group of individuals.

Box 4

BRIEF DEFINITIONS OF KEY TERMS (CONTINUED)

Reliability. Indicators of reliability tell how dependable an assessment or screening tool is for the purpose it is used. Reliable tools are stable over time and include items that measure the same thing in different ways. For tools that require standardized observation (for example, child care quality observations or ratings of children's behavior), the scores obtained by two different, well-trained observers must be similar to be considered reliable. Statistical measures of reliability are typically reported as correlation coefficients, which range from 0 to 1.0, with a higher value reflecting greater reliability. Many researchers and test developers require that assessment and screening tools have reliability values of 0.7 or higher. For our summary descriptors, we adopted a criterion of 0.65, which reflects a rule of thumb commonly used in the field. Typical indicators of reliability include measures of consistency of results and stability over time:

- **Internal consistency.** If the individual items in an instrument tool measure the same thing (for example, they all assess motor ability or language development), the measure is considered to be internally consistent. One measure of internal consistency is split-half reliability. To demonstrate this, test developers and researchers test a group of individuals, then split the test items in half, usually by grouping the odd- and even-numbered items. If the two groupings of the test items are highly correlated with each other, the split-half reliability is considered to be acceptable. Another measure of internal consistency reliability is based on the correlations among all of the individual test items. This index of internal consistency is called Cronbach's alpha (named after the researcher who developed the statistical formula)
- **Stability.** By this measure, an assessment is reliable to the extent the procedure yields the same result on two different occasions. Test-retest reliability involves

testing the same group of individuals at least twice, with a relatively short interval between assessments, usually no longer than a few days or weeks apart. The higher the test-retest reliability, the more stable the assessment tool is considered to be. Longer periods between administrations of the same assessment will reduce the reliability, partly because the individual's situation (for example, skill) can be expected to change. Some assessment tools have two versions of the same test so that the same skills or behaviors can be assessed a second or third time (as in a pre-post or longitudinal study). In such cases, test developers include information on alternate form reliability. To demonstrate that both forms of the test are essentially equivalent, a random half of a large group of individuals is given one form of the test and the other half is given the other form. Alternate form reliability is demonstrated if the scores of the two groups are highly correlated.

- **Reliability of administration.** Another reliability consideration applies to assessment tools that require an observer to score a child's or parent's behavior or complete a rating or checklist describing the behavior observed. To use such assessments in evaluation, researchers and test developers want to be sure that these ratings can be made consistently. One index of consistency is the extent to which two trained observers obtain the same scores when they do their observations at the same time, although independently. This index is referred to as inter-rater reliability. It is usually reported either as the correlation between the scores or ratings obtained by the two observers or as the percentage of items on which the two agree.

Representativeness of Norming Sample.

Standardized screening and assessment tools provide information about how the children and parents in your program are doing compared to the group (or sample) of individuals the test developers or researchers included in their norming group. Knowing

Box 4

BRIEF DEFINITIONS OF KEY TERMS (CONTINUED)

whether the norming sample was nationally representative or representative of the children or parents in your program is important in deciding whether to use a screening or assessment tool. Most test authors include this information in their manuals. In general, it is better if the norming sample includes individuals of the same age group that you will be assessing, as well as geographic and racial/ethnic diversity, so that the assessment results will be relevant to the families in your program.

Validity. Indicators of a screening or assessment tool's validity provide information about whether the tool measures what it is supposed to for the purpose it is being used. Several types of validity are commonly used:

- **Content validity.** This indicator of validity provides information about whether the screening or assessment tool includes items that are a good representation of the area the tool is supposed to measure. There are no statistics associated with content validity. Instead, it is based on professional judgment from reviews of the items to verify that what they are measuring represents the domain of development that the developer intended them to measure and that they provide variety and a range of difficulty. A good manual will include a description of the procedures followed in ensuring that the content is appropriate and representative.
- **Criterion-related validity.** Criterion-related validity indicates how well performance on the screening or assessment tool compares with a criterion, or an independent measure of what the assessment is designed to predict. The criterion measure can be obtained at about the same time or after some interval:
 - To establish **concurrent validity**, test developers and researchers administer the new screening or assessment

tool as well as a similar, established tool to the same individuals within a few hours or days. If the correlation between the two measures is high, concurrent validity is established. Strict interpretations require concurrent validity to reach levels of .70 or higher, but as a rule of thumb, many researchers accept .50 or higher as acceptable. Sometimes concurrent validity is expressed in terms of percent agreement between the two measures. In this compendium, we consider 80 percent agreement or higher as acceptable.

- To establish **predictive validity**, researchers and test developers determine whether the screening or assessment tool conducted at one time point with a group of individuals is correlated with later functioning (these studies are often conducted over two to five years or more). If the correlation between the two measures obtained across the time interval is high, predictive validity is established. If, for example, a measure of vocabulary at age 3 is highly correlated with a test of reading ability in second grade, the vocabulary test could be said to have predictive validity. In some cases, researchers use other activities or events as the criterion, rather than another assessment. For example, predictive validity might be established by correlating age 3 vocabulary with children's second-grade language report card grades. In general, the younger the child being assessed, the poorer the predictive validity. There is a long history of poor predictive validity among infant tests, with almost none meeting high levels of validity, such as .80. Researchers have advanced many explanations for this, including the important contributions of the different environments to which children are exposed. Because we know the predictive validity of infant and toddler assessment tools is low, in this compendium, we consider a correlation of .40 to be adequate for establishing predictive validity.

This discussion is important for interpreting scores from standardized instruments. Scores from other instruments can also be interpreted meaningfully if you can compare the performance of children or parents across two points in time (such as comparing scores at the beginning and end of their program experience).

Box 4

Scoring. Alone, the scores from screening and assessment instruments (raw scores) have limited value. It is only when they are compared against a similar group (or norming sample) of children with known characteristics that a child's score becomes meaningful. Because of this, instrument developers often provide the user with tables for converting raw scores into scores that are normed to a comparison sample. Below are some of the more frequently used normative scores:

- **Percentile rank.** The percentile rank indicates a score's relative ranking, in units 0 to 100, to other scores in the norming sample. A child whose score is at the 65th percentile has scored higher than 65 percent of the children in the norming sample. However, percentiles are not easily comparable to each other because the raw score difference between percentiles will vary depending on the percentiles' location. The raw score differences between percentiles at the extreme ends of the percentile distribution are larger than raw score differences in the middle of the percentile distribution.
- **Stanine score.** Like percentile ranks, stanine scores provide information on children's performance relative to children in the norming sample, but without the restriction on comparing scores. Stanines divide the normal curve into nine intervals, with the lowest scores falling into the first stanine, the highest scores falling into the ninth stanine, and the fifth stanine straddling the midpoint of the distribution. Except for the two extreme stanines (the first and the ninth), each stanine is one-half of a standard deviation unit, and equal differences between two pairs of stanines represent equal differences in performance. A disadvantage of stanine scores is that they magnify small differences between raw scores that fall on either side of a point separating adjacent stanines.
- **Standardized score.** Standardized scores express the difference between a raw score and the mean score in standard deviation units. Standard scores have the properties of the normal curve and maintain the absolute differences between the raw scores. Thus, the difference in performance between standard scores of 85 and 90 is the same as the difference between standard scores of 55 and 60. Three types of standard scores are often used: T-scores, quotients, and normal curve equivalents (NCEs). T-scores have a mean of 50 and a standard deviation of 10, while quotients have a mean of 100 and a standard deviation of 15, and NCEs have a mean of 50 and a standard deviation of 21.06. Most tests of cognitive abilities have a mean of 100 and a standard deviation of 15. For most standardized tests, we consider scores within 30 points of the mean (from 70 to 130) to be in the "normal" range.
- **Age-equivalent scores.** An age-equivalent score is the average raw score of children at that age in the norming sample. The age-equivalent score corresponding to a child's raw score provides information on the child's level of performance in terms of the age at which that level of performance could be expected, based on the performance of children in the norming sample.
- **Sensitivity** is a measure of an instrument's ability to correctly identify persons with the disorder as having the disorder.
- **Specificity** is a measure of an instrument's ability to identify persons who do not have the disorder as not having the disorder.

² A standard deviation is a measure of the score's dispersion or variability in a sample. The proportion of scores within a standard deviation unit of the mean score is known. For example, in a normal distribution, 68 percent of all the scores fall between one standard deviation below and one standard deviation above the mean. Thus, scores expressed in standard deviation units enable the user to understand how a child has performed relative to other children in the sample.

instrument's norming sample, using an unofficial translation or directly translating the instrument into another language will result in scores that may not be comparable to the norming sample scores. According to the strictest standards, such scores are not valid.

- **Type of Instrument.** We categorized the child and parent instruments as one of three types: (1) direct child or parent instruments, in which a trained individual works one-on-one with the child or parent to administer the instrument; (2) observation, in which a trained individual observes the child or parent and either rates or scores the behaviors of interest; and (3) parent report or self-report, in which the parent reports about the child or himself or herself. These basic categories apply to most of the other areas we reviewed as well, such as quality of program services. As needed, we used different descriptors to make our meaning as clear as possible.
- **Age Range and Administration Interval.** We have included the age range for which the instrument is appropriate, as well as the recommended time between administrations of the instrument, if given. Some instruments are designed to be administered at regular intervals, and that information is also included.
- **Personnel, Training, Administration, and Scoring Requirements.** We described whether the instrument requires administration by a consultant or expert with clinical training, a highly trained program staff member, or a clerical program staff member. We included an estimate of how much time a person at the level required would need to learn, conduct, and score the

instrument. Some of the authors and publishers suggest that trainees have an administration reviewed by an experienced assessor. If so, we also included this requirement. Some of the authors and publishers offer group training on the use of their instruments, and we included that information and the cost of the training, if it is available.

- **Summary.** We chose five key features of the instruments to include in the summary table. Each feature has descriptors numbered from 1 to 3. A descriptor of 1 indicates a lack of information or lower-level performance on the feature, a descriptor of 3 indicates a higher-level of performance, and 2 is intermediate. We include this summary section to help you compare the features of the instruments, but do not consider this information as a recommendation of one instrument or another. Only you and your staff can decide which features are most important to you. The purposes of your screening and assessment must guide your choices about which instruments to use. The features we include in the summary section are:

- **Initial material cost:** 1 (under \$100), 2 (\$100 to \$200), 3 (more than \$200).

- **Reliability:** 1 (none described); 2 (all or mostly under .65); 3 (all or mostly .65 or higher). See Box 4 for a brief definition of the various types of reliability. We chose these groupings based on the prevalent rule of thumb researchers and assessment developers use. Other things being equal, the higher the reliability is, the better the instrument is.

- **Validity:** 1 (none described); 2 (all or mostly under .5 for concurrent; all or mostly under .4 for predictive); 3 (all or mostly .5 or higher for concurrent; all or mostly .4 or higher for predictive). See Box 4 for a brief definition of the various types of validity. We chose these groupings based on the prevalent rules of thumb researchers and instrument developers use. Generally, the higher the validity is, the better. It is especially challenging to create instruments for infants and toddlers that strongly predict how the children will do as preschoolers. Therefore, the grouping for predictive validity reflects a less stringent criterion for the highest grouping.

- **Norming sample characteristics:** 1 (none described); 2 (older than 15 years, not nationally representative or representative of the low-income population enrolled by Head Start programs serving infants and toddlers); 3 (normed within past 15 years, nationally representative or representative of the low-income population enrolled by Head Start programs serving infants and toddlers). See Box 4 for a brief definition of representativeness of the norming sample. This section also includes information on the date that the norming sample was obtained. The more time that has elapsed since the norming sample was obtained, the less likely it is to be representative. Many authors/publishers re-norm their assessments every 10 to 12 years to keep them up-to-date. We chose 15 years as the critical time here.

- **Ease of administration and scoring:** 1 (not described); 2 (self-administered or administered and scored by someone with basic clerical skills); 3 (administered and scored by a highly trained indi-

vidual). The administration and scoring requirements for each instrument vary and these descriptors help you determine what is involved for these steps.

The other topics included for each instrument are:

- **Description.** This section provides an overview of what the instrument was designed to measure, the age range of individuals it may be used with, the number of items, how it is administered, and what types of information can be derived (including any scores and subscale scores).

- **Uses of Information.** To help you match your intended purposes for an instrument with the results, we included a summary of how the information that comes from an instrument may be used. Some of the instruments are clearly designed for screening children, some for in-depth assessment, some for allowing comparisons to a national norming sample, some for parent education, and some for feeding back into individual intervention planning and continuous program improvement.

- **Reliability.** Indicators of an instrument's reliability help determine whether an instrument is dependable. For example, a dependable instrument is also stable, and the results would be similar if the instrument was administered to the same individual several times in a short period. Box 4 summarizes key information about what to look for in reports of an instrument's reliability. The types of reliability summarized in the resource document entries include:

- **Measures of internal consistency** (split-half reliability, internal consistency reliability) that indicate the extent to which the items in the instru-

ment “hang together” and tell a coherent story about the child or adult’s functioning

- **Measures of stability** (test-retest reliability, alternate form reliability) that indicate the extent to which the instrument yields the same results when used at different times or using a different form of the instrument (for those that have multiple forms)

- **Measures of the reliability of administration** (inter-rater reliability) that indicate the extent to which two different observers or instrument administrators would interpret and record the information in the same way

• **Validity.** Indicators of an instrument’s validity help determine whether the instrument really measures what it is supposed to for the purpose it is being used. For example, if an instrument is supposed to provide an estimate of a toddler’s language production, how the child performs on the instrument should be similar to how the child performs on another established instrument of language production. We summarize key information about what to look for in reports of an assessment’s validity in Box 4. The types of validity summarized in the resource document entries include:

- **Content validity**, which relies on expert judgment to determine that an instrument actually measures what it is intended to measure

- **Criterion-related validity**, including concurrent validity, which indicates how well the instrument results relate to other information collected at the same time, and predictive validity, which indicates the extent to which the instrument results are

related to later functioning

• **Method of Scoring.** Child screening and assessment instruments may be scored using a simple pass/fail point system, or they may use a broader range of response categories, such as whether the child usually exhibits a particular behavior, is just starting to show the behavior, or does not yet display the behavior. In this section, we summarize the response categories used in the instrument and the types of scores it is possible to compute.

• **Interpretability.** Many instrument authors and publishers provide information about how to interpret what a score or range of scores means as to whether the child is functioning at the level expected for his or her age or whether additional information may be needed. These guidelines are helpful in making sense out of the results. In this section, we summarize what is available to help you interpret the information that comes from each instrument.

• **Training Support.** In this section, we summarize what training in the use of the instrument the authors and publishers recommend. We also describe training materials, products, or sessions available. Some authors and publishers include a lot of information about how to prepare to administer their instruments, while others provide little. Some provide training videotapes or exercises as part of the purchase of the instrument. In this section, we summarize what the authors and publishers include to help you identify who needs to administer the instrument and the resources available for training them.

- ***Adaptations/Special Instructions for Individuals with Disabilities.*** Some instruments are designed specifically to assess the abilities or performance of individuals with disabilities, but most are not. In this section, we describe adaptations or instructions the authors or publishers included for working with people with disabilities.
- ***Report Preparation Support.*** Some instruments include summary sheets or software to help you prepare individual reports based on the results. These reports may be designed to help you customize the program for a given child or parent or to help you share information with parents. Some instruments also include recommendations on how to present reports to parents.

- ***References.*** In this section, we give the full citations for the instruments, manuals, and other sources of information we used to complete each entry. We also include citations for any other materials the authors/publishers make available about the instrument, such as training videotapes and computer scoring programs.

The entries are organized alphabetically in three groups: (1) measures of child development; (2) measures of parenting, the home environment, and family well-being; and (3) measures of program implementation and quality. In front of each group of entries is a summary table that lists the instruments profiled in that section and summarizes their main features.



References



Administration for Children and Families. *Information Memorandum: Using Child Outcomes in Program Self-Assessment*. Publication no. ACYF-IM-HS-00-18. Washington, DC: U.S. Department of Health and Human Services, August 2000.

Administration for Children and Families. *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Volume I: Technical Report*. Washington, DC: U.S. Department of Health and Human Services, 2002.

Administration on Children, Youth and Families. *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report*. Washington, DC: U.S. Department of Health and Human Services, 2001a.

Administration on Children, Youth and Families. *Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report*. Washington, DC: U.S. Department of Health and Human Services, 2001b.

Administration on Children, Youth and

Families. *Leading the Way: Characteristics and Early Experiences of Selected Early Head Start Programs. Volume I: Cross-Site Perspectives.* Washington, DC: U.S. Department of Health and Human Services, 1999.

Bowman, Barbara T., M. Suzanne Donovan, and M. Susan Burns. *Eager to Learn: Educating Our Preschoolers.* Washington, DC: National Academy Press, 2000, pp. 233-260.

Early Head Start National Resource Center. *Developmental Screening, Assessment, and Evaluation: Key Elements for Individualizing Curricula in Early Head Start Programs.* Technical Assistance Paper #4. Washington, DC: ZERO TO THREE, undated. Available at <http://www.ehsnrc.org/pdffiles/FinalTAP.pdf>

Gilliam, Walter S., and Linda C. Mayes. "Clinical Assessment of Infants and Toddlers." In *Child and Adolescent Psychiatry: A Comprehensive Textbook (3rd Edition)*, edited by M. Lewis. Baltimore, MD: Lippincott, Williams, and Wilkins, 2002.

Martin, Sue. *Take a Look: Observation and Portfolio Assessment in Early Childhood (2nd Edition).* Reading, MA: Addison-Wesley Publishing Company, 1999.

Meisels, Samuel J., and Sally Provence. *Screening and Assessment: Guidelines for Identifying Young Disabled and Developmentally Vulnerable Children and Their Families.* Washington, DC: ZERO TO THREE, 1992.

Shepard, Lorrie, Sharon Lynn Kagan, and Emily Wurtz (eds.). *Principles and Recommendations for Early Childhood Assessments.* Submitted to the National Education Goals Panel by the Goal 1 Early Childhood Assessments Resource Group, January 1998.

U.S. Department of Health and Human Services. "Head Start Program: Final Rule." *Federal Register*, vol. 61, no. 215, November 5, 1996, pp. 57 186-57 2278.

U.S. Department of Health and Human Services. *The Statement of the Advisory Committee on Services for Families with Infants and Toddlers.* Washington, DC: U.S. Department of Health and Human Services, September 1994.

Wolverton, E. Dollie. "An Introduction to Developmental Screening in the Education Component." *National Head Start Bulletin*, no. 43, January/February 1993, pp. 3-4.



WORKSHEET TO HELP DEVELOP A COMPREHENSIVE PLAN FOR GATHERING AND ANALYZING DATA

Appendix A



SUMMARY OF INFORMATION COLLECTED AND SCREENING/ASSESSMENT TOOLS USED BY YOUR HEAD START PROGRAM SERVING INFANTS AND TODDLERS

Head Start programs serving pregnant women and families with infants and toddlers collect information for various purposes, using a variety of assessment tools or instruments. The table below provides a useful format for summarizing your program's current or planned use of assessment tools and instruments and their purposes.

Instrument/Tool	When Administered:						How Used:				
	At Enrollment	Scheduled Times After Enrollment	At Specific Ages	Under Certain Circumstances	Varies	Other	To Plan Individual Services	To Create Aggregate Reports For:			
							Program Planning	Funders	PIR	Other	
Assessing Family Social Service Needs											
Assessing Quality of Program Services											
Tracking Service Delivery											

Instrument/Tool	When Administered:						How Used:				
	At Enrollment	Scheduled Times After Enrollment	At Specific Ages	Under Certain Circumstances	Varies	Other	To Plan Individual Services	To Create Aggregate Reports For:			
								Program Planning	Funders	PIR	Other

Assessing Family Engagement/Satisfaction

Assessing Overall Program Performance and/or Adherence to the Performance Standards

Any Other Data Collection/Assessments



PERFORMANCE MEASURES FRAMEWORK FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

Appendix B



As a national laboratory for early childhood education, Head Start has long emphasized continuous program improvement and has been a leader in developing outcomes-oriented accountability. Head Start began developing program performance measures in 1995, and in 1997 the Family and Child Experiences Survey (FACES) was launched to collect data on the performance indicators, in part to be responsive to the Government Performance and Results Act (GPRA) requirements. Following the re-authorization of Head Start in 1998, Head Start programs were required to include child outcomes in their self-assessment process by 2003.

Although Early Head Start programs have not been required to report child outcomes, many have started to define and measure outcomes, for several reasons. Some programs operate within Head Start programs and are doing so in conjunction with outcomes assessment in Head Start. Some programs are responding to the requirements of other funders. Yet others are acting out of a desire to improve their services to families with infants and toddlers. In Spring 2001, the Early Head Start Technical Work Group recom-

mended that the Head Start Bureau move forward to develop performance measures for Head Start programs serving infants and toddlers to support programs' efforts. The framework presented in this document resulted from an iterative process during which a wide range of stakeholders were consulted.

The Head Start performance measurement framework needed to be modified for infants and toddlers in order to be appropriate and useful for home-based, center-based, combination, and locally-designed programs, and it needed to take into account the full range of development over a much longer period, from pregnancy to age 3.

The recent completion of the national Early Head Start Research and Evaluation Project also makes the time ripe for developing a new performance measures framework. The pattern of program impacts by age 3 and variations in impacts by program approach (center-based, home-based, and mixed approaches) and key aspects of program implementation provide insights into ways that the Head Start framework must be adapted for programs serving families with infants and toddlers.

Building the Pyramid

The conceptual model underlying performance measures for Head Start programs serving infants and toddlers is based on the Head Start model (ACYF 2001), which was modified to reflect services for infants and toddlers.

The Statement of the Advisory Committee on Services for Families with Infants and Toddlers

recommended four cornerstones for Early Head Start: child development, family development, staff development, and community building (Administration on Children, Youth and Families 1994). Thus, the adapted pyramid rests on these cornerstones, and the elements of management systems related to each cornerstone.

The overall goal of the base level of the pyramid is "ensuring well-managed programs that involve parents in decision-making." This goal links the key elements in management systems that correspond to the four cornerstones.

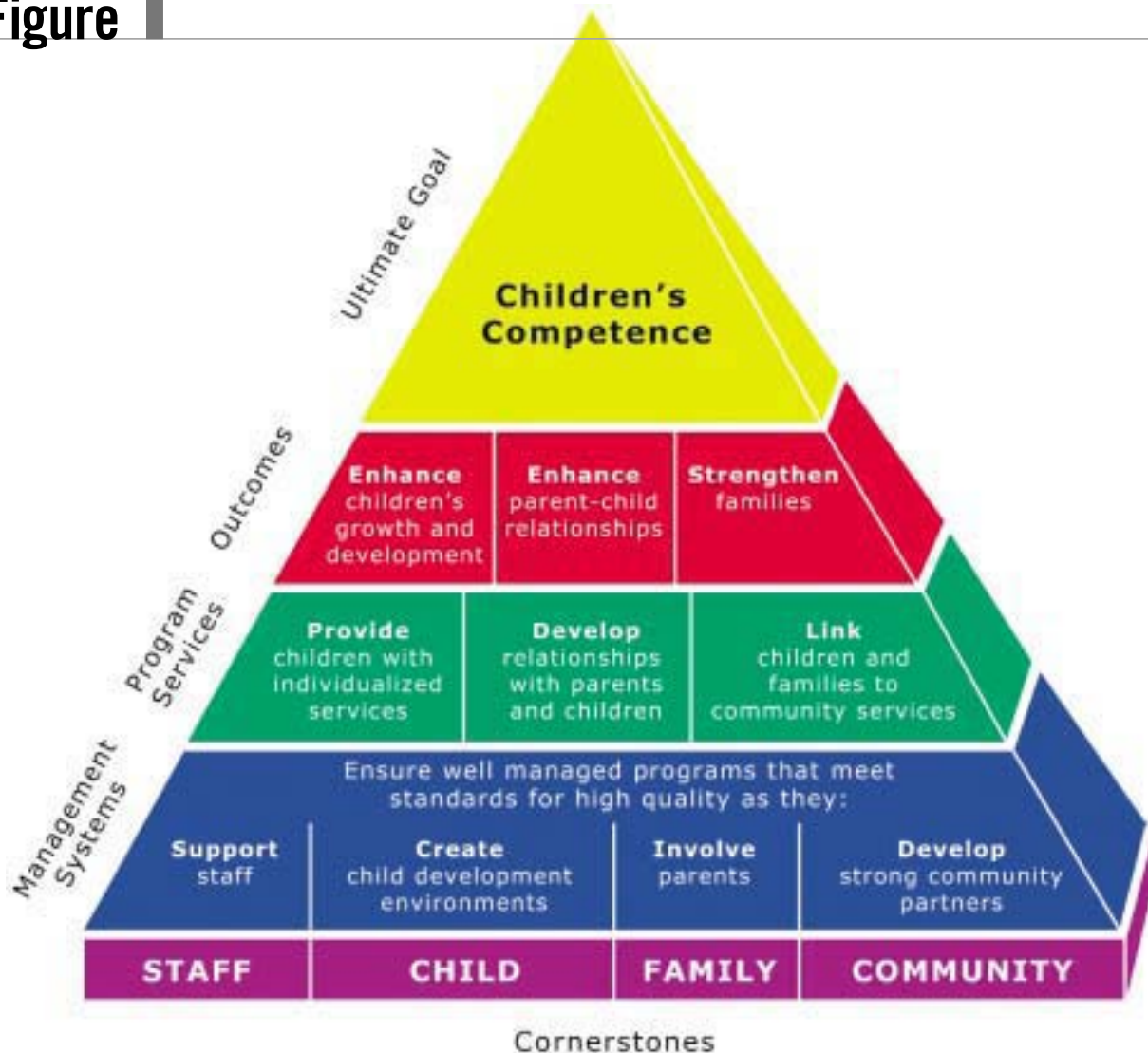
The services level of the pyramid has three main objectives: providing children with services, linking children and families to needed community services, and developing relationships with parents and children.

Similarly, on the level of child and family outcomes, there are three main objectives: enhancing children's growth and development, strengthening families as primary nurturers of their children, and enhancing parent-child relationships. Because it is through these relationships that strengthening families can support enhanced child outcomes and the ultimate goal of enhanced children's competence, the objective of enhancing parent-child relationships was placed between the other two objectives on the outcomes level.

Head Start programs serving pregnant women and families with infants and toddlers may provide center-based services, home-based services, or some combination of these, reflecting different family needs, different theories of change, and different emphases on certain services and outcomes.

Figure 1

Conceptual Framework for Programs Serving Infants and Toddlers



The pyramid for these programs was designed to encompass these key variations. When providing center-based services, programs emphasize the left side of the pyramid—they aim to improve children’s competence mainly by providing direct center-based services to children to enhance their growth and development (and also provide services to strengthen families and improve parent-child relationships). When providing home-based services, programs emphasize the right side of the pyramid—they aim to improve children’s compe-

tence mainly by providing services to strengthen families and improve parent-child relationships (and also work with children directly during home visits and in parent-child group socialization activities). These variations in program emphases were reflected in the differences in impacts found in the national Early Head Start Research and Evaluation project (ACF 2002).

Performance Measures Underlying the Main Objectives

Underlying the main objective in each block in the pyramid are performance measures that represent key program goals under that objective (Figure 2). The program performance measures for Head Start programs serving infants and toddlers include selected performance measures from the Head Start framework, as well as new performance measures that reflect the special features of programs serving infants and toddlers, variations in program approaches, and lessons from the national evaluation and training and technical assistance efforts. The performance measures also reflect key requirements in the Head Start Program Performance Standards as well as elements from the Advisory Committee on Services for Families with Infants and Toddlers.

Using The Program Performance Measures Framework

The program performance measures framework for Head Start programs serving pregnant women and families with infants and toddlers is designed to support continuous program improvement efforts of individual programs, as well as those at the regional and national levels. At the individual program level, the framework, along with the program's specific theory of change, can guide the development of plans for data collection to provide the program staff with important information on program strengths and weaknesses. This information can help focus program improvement efforts on areas where improvement may be needed most and in ways that may be most beneficial to the program.

Information on strengths can be used to highlight program accomplishments and build support for the program among funders and key stakeholders. *A Resources for Measuring Services and Outcomes* is designed to support programs in these activities by providing information on creating an integrated plan and selecting measures.

References

Administration for Children and Families. Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Washington, DC: U.S. Department of Health and Human Services, June 2002.

Administration on Children, Youth and Families. First Progress Report on the Head Start Program Performance Measures. Washington, DC: ACYF, May 1997.

Administration on Children, Youth and Families. The Statement of the Advisory Committee on Services for Families with Infants and Toddlers. Washington, DC: U.S. Department of Health and Human Services, 1994.

Administration on Children, Youth and Families. Head Start Program Performance Measures: Second Progress Report. Washington, DC: U.S. Department of Health and Human Services, June 1998.

Administration on Children, Youth and Families. Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report. Washington, DC: U.S. Department of Health and Human Services, January 2001.

Figure 2

PROGRAM PERFORMANCE MEASURES FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

FOUNDATION: PROGRAM MANAGEMENT SYSTEMS

Management systems (including program governance, planning, communication, record-keeping and reporting, ongoing monitoring, self-assessment, human resources, and fiscal management) ensure well-managed programs that meet standards for high quality as they create nurturing child development environments, enhance child outcomes and promote positive parent-child interactions.

1. Programs comply with Head Start regulations.
2. Programs are well-managed operationally and financially
3. Programs design and implement services to be responsive to the needs of families in the community.
4. Programs conduct self-assessments that are used for continuous program improvement.

Management Systems ensure well-managed programs with integrated systems to support staff in working effectively with parents and children.

1. Programs employ qualified staff with the skills necessary to provide high-quality services.
2. Programs support ongoing staff development, training, and mentoring.
3. Programs support staff activities through ongoing reflective supervision.
4. Programs promote staff retention and continuity.

Ensure well-managed programs that meet standards for high quality as they develop collaborative relationships with community partners.

1. Programs form partnerships with other community programs and organizations to support an integrated community-wide response to the needs of families with young children.
2. Programs form partnerships and coordinate services with local Part C agencies.
3. Programs form partnerships and coordinate services with community child care providers to meet the needs of families and enhance the quality of local child care services through the sharing of resources, training, and knowledge.
4. Programs form partnerships and coordinate services with local health agencies and health care providers to meet the health-related needs of families.

Management Systems ensure well-managed programs that meet standards for high quality as they involve parents.

1. Parents are involved actively in program planning and decision making.
2. Programs encourage and support fathers' involvement in program planning, decision-making, and activities.

PROGRAM SERVICES

Programs develop individualized family partnership agreements with families and link children and families to comprehensive community services in order to meet their personal goals and self-sufficiency needs.

1. Programs work collaboratively with families to identify their goals, strengths, and needed services, and offer them opportunities to develop

and implement individualized family partnership reements that take into account other family plans.

2. Programs link parents with social service agencies to obtain needed services.
3. Programs link parents with educational and employment agencies to obtain needed services.
4. Programs link parents with physical and mental health care prevention and treatment services to obtain needed care.
5. Programs link parents with needed prenatal care and education services.
6. Programs help parents secure high-quality child care in order to work, attend school, or gain employment training.
7. Programs help parents and children make a smooth transition to Head Start or other pre-school program.

Program staff develop responsive and caring relationships with parents and children

1. Staff form respectful and supportive relationships with parents through all aspects of service delivery.
2. Staff form nurturing relationships with children in group-care settings or during home visits.
3. Programs support and honor the home cultures and languages of families.

Programs provide children with age-appropriate curricular experiences and individualized edu-

cational, health, and nutritional services

1. Programs provide developmentally enriching educational environments in group-care settings and developmentally enriching parenting and child development services during home visits and group socializations.
2. Programs link children with needed medical, dental, and mental health services.
3. Programs link pregnant women with comprehensive prenatal health care and education.
4. Programs provide children in group-care settings meals and snacks that meet their daily nutritional needs, and parents receiving home-based services receive information about meeting their children's nutritional needs
5. Programs provide individualized services for parents and children, including children with disabilities.

EXPECTED OUTCOMES FOR CHILDREN AND FAMILIES

Programs strengthen families as the primary nurturers of their children

1. Parents demonstrate increased knowledge of child development and awareness of their children's developmental progress
2. Parents enhance their self-concept and emotional well-being and experience less parenting stress.
3. Parents make progress toward their educational, literacy, and employment goals.
4. Adult family members strengthen their relation-

ships and work together in caring for children.

Programs support and enhance parent-child relationships

1. Parents demonstrate more sensitivity and responsiveness in interactions with their children.
2. Parents spend more time with their children in activities that stimulate their children's development, such as reading to their children
3. Parents provide home environments and experiences that are more supportive of their children's development.

Programs enhance children's growth and development

1. Children demonstrate improved communication, language, and emergent literacy skills.
2. Children demonstrate improved general cognitive skills.
3. Children demonstrate improved positive approaches toward learning, including improved attention skills.
4. Children demonstrate improved social behavior, emotion regulation, and emotional well-being.
5. Children demonstrate improved physical health and development.

ULTIMATE GOAL: CHILDREN'S COMPETENCE



LIST OF MEASURES USED IN THE NATIONAL EARLY HEAD
START RESEARCH AND EVALUATION PROJECT

Appendix C



Measure	Type*	Child's Age			Construct
		14-mo	24-mo	36-mo	
Achenbach System of Empirically-Based Assessment, Child Behavior Checklist, Aggressive subscale (Achenbach and Rescorla 2000)**	CH		X	X	Aggressive behavior problems
Arnett Scale of Caregiver Behavior (Arnett, 1989)**	CL	X	X	X	Caregiver's sensitivity, harshness, and detachment
Bayley Behavioral Rating Scale (Bayley 1993)**	CH	X	X	X	Engagement of examiner, emotional regulation during assessment
Bayley Scales of Infant Development-II, Mental Development Index (Bayley 1993)**	CH	X	X	X	Cognitive development
Beliefs Regarding Talking and Reading scale (Luster, Rhoades and Haas, 1989)	F		X		Parent's beliefs about talking and reading to children
Center for Epidemiological Studies - Depression Scale (CES-D; Radloff 1977)**	F	X			Parent's depressive symptoms
Center for Epidemiological Studies - Depression Scale, short form (CES-D; Radloff 1977; Ross, Mirowsky and Huber 1983)	F			X	Parent's depressive symptoms
Child-Caregiver Observation System (C-COS; Boller, Sprachman, and the Early Head Start Research Consortium 1998)	CL		X	X	Quality, frequency, and types of provider-child interactions
Composite International Diagnostic Inventories (CIDI) - Short Form, Major Depression, Generalized Anxiety Disorder, Alcohol Dependence, Drug Dependence (Nelson, Kessler, and Mroczek 1998 and World Health Organization 2002)**	F		X		Parent's major depression, generalized anxiety disorder, alcohol abuse, and substance abuse
EAS Temperament Survey for Children (Buss and Plomin, 1984), selected items	CH	X			Temperament (emotionality and sociability)
Early Childhood Environmental Rating Scale-Revised (ECERS-R; Harms, Clifford, and Cryer 1998)**	CL			X	Quality of the caregiving environment in center-based care

Measure	Type*	Child's Age			Construct
		14-mo	24-mo	36-mo	
Family Day Care Rating Scale (FDCRS; Harms and Clifford 1989)**	CL	X	X	X	Quality of the caregiving environment in home-based care
Family Environment Scale, Conflict subscale (Moos and Moos, 2002)**	F	X	X	X	Family conflict
High Chair (Frustration) Assessment (Shaw et al. 1994; McHale et al. 1999)	F	X			Parent's Sensitivity, Detachment, Negative Regard/Hostility, Positive Regard, Distance/Approach
High Chair (Frustration) Assessment (Shaw et al. 1994; McHale et al. 1999)	CH	X			Child Distress
Home Observation for Measurement of the Environment (HOME), Infant/Toddler Form (Caldwell and Bradley, 2003)**	F	X	X		Quality of home environment, Reading and language activities, Emotional Responsivity, Support of Cognitive, Language and Literacy Environment, Maternal Verbal-Social Skills, Absence of Punitive Interactions
Home Observation for Measurement of the Environment (HOME), Preschool Form, NLSY version, selected items (Caldwell and Bradley, 2003 and Center for Human Resources Research, 2000)	F			X	Quality of home environment, Reading and language activities, Warmth, Support of Language and Learning, Internal Physical Environment, Harshness
Discipline Vignettes	F	X	X	X	Parental use of reasoning and developmentally appropriate approaches in parent-child conflict situations
Infant/Toddler Environment Rating Scale (ITERS, Harms, Cryer, and Clifford 1990)**	CL	X	X		Quality of the caregiving environment in center-based care
Knowledge of Infant Development Inventory (KIDI, MacPhee, 1981), selected items**	F	X	X		Parent's knowledge of childrearing practices, developmental processes, and infant developmental norms and milestones
MacArthur Communicative Development Inventories (CDI), Vocabulary Short Form, Level I; Part II, Section A First Communicative Gestures; and Part II, Section B Games and Routines (MacArthur CDI Advisory Board 1997; and Fenson et al. 2000a and b)**	CH	X			Receptive vocabulary, productive vocabulary, use of gestures to communicate, parent-child communicative activities
MacArthur Communicative Development Inventories (CDI), Vocabulary Short Form, Level II; Part II, Question on Combining Words; and Section E. Complexity (MacArthur CDI Advisory Board 1997; Fenson et al. 2000a and b)**	CH		X		Productive vocabulary, complexity of speech

Measure	Child's Age				Construct
	Type*	14-mo	24-mo	36-mo	
Nursing Child Assessment Satellite Training (NCAST) Parent-Child Interaction Program, Teaching Scales (Barnard 1994; Sumner and Spietz 1994)**	F		X		Parent total score
Nursing Child Assessment Satellite Training (NCAST) Parent-Child Interaction Program, Teaching Scales (Barnard 1994; Sumner and Spietz 1994)**	CH		X		Child total score
Child's Health Status (National Health Interview Study)	CH	X	X	X	Child's health status
Parent-Caregiver Relationship Scale (selected items) (Elicker, et al. 1996)	F	X	X	X	Parent's relationship with current nonrelative provider
Parent-Caregiver Relationship Scale (selected items) (Elicker, et al., 1996)	CL	X	X	X	Caregiver's relationship with parents
Parental Modernity Scale (selected items) (Schaefer and Edgerton 1985)	F		X		Parent attitudes toward children and childrearing practices (traditional attitudes and practices; progressive attitudes and practices)
Parental Modernity Scale (selected items) (Schaefer and Edgerton 1985)	CL		X	X	Caregiver attitudes toward children and childrearing practices (traditional attitudes and practices; progressive attitudes and practices)
Parenting Stress Index, Short Form (Abidin, 1995), Parental Distress and Parent-Child Dysfunctional Interaction Subscales**	F	X	X	X	Parental distress, Parent-child dysfunctional interaction
Peabody Picture Vocabulary Test-III (PPVT-III; Dunn and Dunn 1997) or Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Padilla, Lugo, and Dunn, 1986)**	CH			X	Receptive Language
Provider Job Commitment (Kontos et al. 1995)	CL	X	X	X	Provider's commitment to child care as a profession
Provider Motivation (Kontos et al. 1995)	CL	X	X	X	Provider's motivation for beginning child care work

Measure	Type*	Child's Age			Construct
		14-mo	24-mo	36-mo	
Provider Views about Training (Kontos et al. 1995)	CL	X	X	X	Provider's views about the usefulness of training
Puzzle Challenge (Problem-Solving) Task (Brady-Smith et al. 2001)	F			X	Parent Supportive Presence, Quality of Assistance, Detachment, Intrusiveness
Puzzle Challenge (Problem-Solving) Task (Brady-Smith et al. 2001)	CH			X	Child Engagement of Parent, Persistence, Frustration with Task
Three-Bag Assessment (Semi-structured Play) (NICHD Study of Early Child Care 1992; Ware et al. 1998; Brady-Smith et al. 1999 and 2000)	F	X	X	X	Parent Supportiveness, Sensitivity, Positive Regard, Stimulation of Cognitive Development, Detachment, Intrusiveness, Negative Regard
Three-Bag Assessment (Semi-structured Play) (NICHD Study of Early Child Care 1992; Ware et al. 1998; Brady-Smith et al. 1999 and 2000)	CH	X	X	X	Child Engagement of Parent; Sustained Attention to Objects, Negativity Toward Parent
Student-Teacher Relationship Scale (Conflict and Closeness subscales) (Pianta 1992)	CH			X	Conflict with child; Closeness toward child (provider report)
Mastery Scale (Pearlin and Schooler 1978)	F	X			Parent's self-efficacy
Woodcock-Munoz Language Survey: Picture Vocabulary Test (Woodcock and Munoz-Sandeval 2001)	F		X		Parental language/cognition

* CH - Child Measure, F - Family/Parent Measure, CL - Classroom Measure, S - Staff Measure

** A profile of this measure is included in this resource document.

Child Development Instruments



Child Development Instruments

Instrument	Screening or Assessment	Domain			Age Range	Assessment Type	Initial Material Cost	Reliability	Validity	Norming Sample	Ease of Administration and Scoring
		C	L	S-E							
Achenbach System of Empirically Based Assessment	A	X	X	X	1.5-5 years	3	2	3	3	3	3
Ages and Stages Questionnaires	S	X	X	X	4-60 months	3	2	3	2	2	2
Ages and Stages Questionnaires:Social-Emotional	S			X	3-66 months	3	2	3	3	2	3
Assessment, Evaluation, and Programming System for Infants and Children	A	X	X	X	0-3 years	2,1,3	2	3	3	1	2
Batelle Development Inventory	A	X	X	X	0-8 years	1,2,3	1	3	3	2	2
Bayley Scale for Infant Development, Second Edition	A	X	X	X	1-42 months	1	3	3	2	3	3
Brigance Diagnostic Inventory of Early Development-Revised	A	X	X	X	0 to 7 years	1,2,3	3	3	2	1	3
Carey Temperament Scales	A			X	0-12 years	3	1	2	1	2	2
Carolina Curriculum for Infants and Toddlers with Special Needs Assessment Log II	A/S	X	X	X	0-2 years	2	1	1	1	1	3
Denver II Development Screening Test	S	X	X	X	0-6 years	1,3	1	3	1	2	3
Developmental Observation Checklist System	S	X	X	X	0-6 years	3	2	3	2	2	3
Developmental Profile II	A	X	X	X	0-7 years	1,3	2	3	2	2	2
Devereux Early Childhood Assessment	A			X	2-5 years	1,2	2	2	1	3	3
Early Coping Inventory	A			X	4-36 months	2	1	3	1	1	3
Early Head Start Evaluation--Parent Interviews and Child Assessments	A/S	X	X	X	14,24,36 months	1,2,3	1	3	1	1	2,3
Early Learning Accomplishment Profile--Revised Edition	A	X	X	X	0-36 months	2	3	3	3	3	3
Eyberg Child Behavior Inventory/Sutter-Eyberg Student Behavior Inventory-Revised	S		X	X	2-16 years	3	NA	3	2	2	3
Functional Emotional Assessment Scale	A		X	X	7-48 months	2,1	1	3	2	1	3
Hawaii Early Learning Profile	A	X	X	X	0-36 months	1	1	1	1	1	3
High/Scope Child Observation Record	A	X	X	X	30-72 months	2	2	3	2	1	2
High/Scope Child Observation Record for Infants and Toddlers	A	X	X	X	1-36 months	2	2	3	3	1	3
Humanics National Infant-Toddler Assessment	A	X	X	X	0-3 years	2	1	1	1	1	2
Infant-Toddler Developmental Assessment	A	X	X	X	0-42 months	2,3	3	3	3	2	3
Infant Toddler Social-Emotional Assessment	S		X	X	1-4 years	3	1	3	3	2	2
Infant Toddler Symptom Checklist	S		X	X	7-30 months	3	1	1	2	1	2
Leiter International Performance Scale-Revised	S/A	X		X	2-20 years	1	3	3	3	3	3
MacArthur Communicative Development Inventories	A		X		8-30 months	3	3	3	3	2	2
Mullen Scales of Early Learning	A	X	X	X	0-68 months	1	3	3	3	3	3
The Ounce Scale	A	X	X	X	0-42 months	2,3	NA	NA	NA	NA	NA
Peabody Picture Vocabulary Test, Third Edition	A		X	X	2.5-90 years	1	2	3	1	3	2
Preschool Language Scale	A		X	X	0-6 years	1	2	3	1	3	3
Receptive-Expressive Emergent Language Test--2nd Ed.	S		X	X	0-3 years	1	1	3	1	2	3
Rossetti Infant-Toddler Language Scale	A		X	X	0-3 years	1,2,3	1	1	1	1	3
Temperament and Atypical Behavior Scale	S/A		X	X	11-71 months	3	1	3	1	1	3
Test de Vocabulario en Imagenes Peabody	A		X		2-18 years	1	2	3	1	2	3
Vineland Adaptive Behavior Scale	A		X	X	0-19 years	3	2	1	3	2	3
Vineland Social-Emotional Early Childhood Scales	A		X	X	0-5 years	3	1	3	3	2	3
Woodcock-Johnson III	A	X	X	X	2-90 years	1	3	3	3	3	3

KEY

- Domains**
- C = Cognitive, problem-solving
 - L = Language, communication
 - S-E = Social, emotional
 - M = Motor, physical
 - O = Other
- Norming sample**
- 1 = None described
 - 2 = Older than 15 years, not nationally representative or representative of EHS population
 - 3 = Normed within past 15 years, nationally representative or representative of EHS population
- Assessment Type**
- 1 = Direct assessment
 - 2 = Observation
 - 3 = Parent/Teacher self report
- Initial Material Cost**
- 1 = Under \$100
 - 2 = \$100 to \$200
 - 3 = More than \$200
- Reliability**
- 1 = None described
 - 2 = Under .65
 - 3 = .65 or higher
- Validity**
- 1 = None described
 - 2 = Under .5 for concurrent; under .4 for predictive
 - 3 = .5 or higher for concurrent; .4 or higher for predictive
- Ease of administration and scoring**
- 1 = Not described
 - 2 = Self-administered or administered and scored by someone with basic clerical skills
 - 3 = Administered and scored by a highly trained individual

The information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

ACHENBACH SYSTEM OF EMPIRICALLY BASED ASSESSMENT (ASEBA), 1999

Authors:

Thomas M. Achenbach and Leslie A. Rescorla

Publisher:

ASEBA

(802) 656-8313 or 656-3456

www.ASEBA.org

Initial Material Cost:

Preschool hand-scoring starter kit: \$174

Representativeness of Norming Sample: The 1999 child behavioral checklist norming sample of 700 children is nationally representative, but restricted to children with no major physical or mental disabilities and English-speaking parents. The 1,192 caregiver-teacher norming sample is not nationally representative—989 caregivers from the 1997 norming sample augmented the 203 preschool caregivers-teachers drawn from the 1999 sample. The language development survey sample consisted of 278 parents from the 1999 sample.

Languages:

English and Spanish (child behavioral checklist only)

Type of Assessment:

Parent and caregiver report

Age Range and Administration Interval: 1.5 to 5 years. No prescribed interval, but routine use is recommended.

Personnel, Training, Administration, and Scoring Requirements:

Respondents should be able to read at the 5th grade level or higher and complete the forms in about 10 to 15 minutes. The authors recommend that a person with graduate training and familiarity with the manual interpret and score the assessment.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher, with some lower)

Validity: 3 (.5 or higher for concurrent, with some lower)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (self-administered but scored by a highly trained individual)

Description: The ASEBA uses information collected from parents and caregivers/teachers to assess the behavioral, emotional, and social functioning (including language development) of young children between the ages of 1.5 and 5 years. To get a better understanding of how the child functions under different conditions, it is recommended that information be collected from more than one adult. The ASEBA consists of two

self-administered reporting forms. The parent report consists of a 99-item child behavioral checklist (CBC) and a language development survey (LDS) that asks parents to provide the child's best multi-word phrases and words the child uses from a list of 310 words. The caregiver/teacher report (CTR) consists of a 99-item checklist similar to the CBC except 17 family-specific items have been replaced with group situation items.

The 99 items in the CBC are organized into seven syndromes and two broader groupings of syndromes, while the 99 items in the CTR are organized into six of the CBC syndromes and the two broader groupings:

- Internalizing
 - Emotionally reactive
 - Anxious/depressive
 - Somatic complaints
 - Withdrawn
- Externalizing
 - Attention problems
 - Aggressive behavior
- Ungrouped (CBC only)
 - Sleep problem

The items are also organized into five DSM (Diagnostic and Statistical Manual of Mental Disorder)-oriented scales.

Uses of Information: The results can be used to structure interviews with parents, identify areas for intervention, and evaluate intervention outcomes.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): the alphas for the CBC scales ranged from .66 to .92 for the syndromes and .63 to .86 for the DSM-oriented scales. The alphas were .89 and .92 for the two broader groupings (internalizing and externalizing syndromes) and .95 for the total score. The alphas for the CTR syndromes ranged from .52 to .96 and for the DSM-oriented scales from .68 to .93. The alphas were .89 and .96 for the internalizing and exter-

nalizing groupings and for the total score, .97. (2) Test-retest reliability, with an eight-day interval between tests: the correlations were .85 and .76 for the CBC and CTR, respectively. Test-retest studies on the LDS reported correlations greater or equal to .90.

Validity: (1) Concurrent validity: The CBC correctly classified 84 percent of a sample of children (some of whom were diagnosed as having emotional/behavioral problems), and the CTR correctly classified 74 percent of the children. Studies reported correlation coefficients between the CBC problem syndromes and the Toddler Behavior Screening Inventory and the Infant-Toddler Social and Emotional Assessment ranging from .48 to .70. In 11 studies that compared parent LDS scores with those obtained by trained examiners using other measures, the correlations between the parent's score and the trained examiner's ranged from .56 to .87. Other studies found the level of LDS agreement with other measures of language development ranged from .47 to .94. (2) Predictive validity: An 11-year longitudinal study found that children identified by the LDS to have language development problems were more likely to have weak verbal skills at age 13.

Method of Scoring: ASEBA can be hand or computer scored. Respondents complete the CBC and CTR by circling one of three responses and the LDS by circling the words the child uses spontaneously. The behavioral raw scores are derived by summing the response item values (0, 1, or 2) for the syndrome scale, syndrome groupings, and total score. The raw score for the language devel-

opment survey is the total number of circled words. The manual provides instructions for converting raw scores into T-scores.

Interpretability: Although the ASEBA provides the user with T-scores to compare a child's performance against other children and the scoring forms classify scores as normal (under 93 percent), borderline (93 to 97 percent), or clinical (over 97 percent), the authors recommend that the results be interpreted by someone with some graduate training.

Training Support: None indicated, however, ASEBA was designed to be easy to use and some support may be available on the internet.

Adaptations/Special Instructions for

Individuals with Disabilities: The manual does not provide details about this, but suggests that persons rating children with disabilities compare the child's behavior to their expectations of a typical same-age child.

Report Preparation Support: The manual shows a report generated by computer scoring software.

References:

Achenbach, Thomas M. and Leslie A. Rescorla. *Manual for the ASEBA Preschool Forms and Profiles*. ASEBA, Burlington, VT, 2000.

AGES AND STAGES QUESTIONNAIRES (ASQ) A PARENT-COMPLETED, CHILD-MONITORING SYSTEM, SECOND EDITION, 1999

Authors:

Jane Squires, LaWanda Potter, and Diane Bricker

Publisher:

Paul H. Brookes Publishing Co.

(800) 638-3775

www.pbrookes.com

Initial Material Cost:

Questionnaires and Users' Guide: \$190 Ages and Stages Questionnaire on a Home Visit (video): \$44

Representativeness of Norming Sample:

Families with children between ages 4 to 36 months from both risk and non-risk populations whose families are educationally, economically, and ethnically diverse (primary sample derived between 1980 and 1988). Normative sample children met the following criteria: (1) no history of developmental or serious health problems; (2) full-term birth; and (3) not placed in a neonatal intensive care unit.

Languages:

English, Spanish, French, and Korean

Type of Assessment:

Parent (or caregiver) report

Age Range and Administration Interval: 4 to 60

months; administered at months 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, 36, 42, 48, 54, and 60 months of age, although users can vary the interval to fit their program or population need.

Personnel, Training, Administration, and Scoring Requirements:

Questionnaires are written at a 6th grade reading level for parents to easily understand and administer. Each questionnaire takes 15 minutes to administer and approximately 1 minute to score. Interpretation of scores requires professionals or trained paraprofessionals.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (under .5 for concurrent and .4 for predictive)

Norming Sample Characteristics: 2 (not nationally representative, older than 15 years)

Ease of Administration and Scoring: 2 (self-administered by parent and scored by trained staff member or parent)

Description: This series of 19 parent-completed questionnaires with 30 developmental items in each questionnaire helps screen infants and young children for developmental delays during their first 5 years. It is completed by parents or caregivers for children 4 to 60 months of age. The questionnaires focus on assessment of five

key developmental areas: communication, gross motor, fine motor, problem solving, and personal-social. The ASQ also includes a section where parents can record general concerns/issues that are not captured in the questionnaire.

Uses of Information: The ASQ provides comprehensive initial screening for developmen-

tal delays, monitoring and identification of areas needing further assessment, and parent education and involvement. It can also be used to monitor at-risk children.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): Communication (.63 to .75); Gross Motor (.53 to .87); Fine Motor (.49 to .79); Problem Solving (.52 to .75); Personal-Social (.52 to .68). (2) Test-retest reliability: percent agreement between administrations was 94 percent. (3) Inter-rater reliability: percent agreement between observers was 94 percent.

Validity: (1) Concurrent validity: percent agreement between the ASQ and other measures (the Revised Gesell and Armatruda Developmental and Neurological Examination and the Bayley Scales of Infant Development) was 84 percent overall and ranged from 76 percent for the 4-month questionnaire to 91 percent for the 36-month questionnaire. (2) Predictive validity: not described.

Method of Scoring: For each item in the ASQ, the parent responds with "yes," "sometimes," or "not yet." These items are then converted to point values and summed. The scorer can then compare the summed total score to established screening cutoff points. Scoring can be done by trained program staff either in their offices or on site during a home visit to give parents immediate feedback). If necessary, parents can also score the questionnaires themselves, using the Information Summary Sheet.

Interpretability: Professionals or paraprofessionals are required to provide feedback to parents

who have completed the questionnaire. There is an Information Summary Sheet intended to assist program staff (or parents) with scoring and provide them with a summary of the child's performance on the questionnaire. The Information Summary Sheets can be kept by program staff as a record of the child's performance on the ASQ so that the actual questionnaires can be returned to the parents for reference.

Training Support: The Users' Guide contains complete instructions for each of the phases of the questionnaire. Other support materials include guidelines for choosing referral criteria, activities sheets that correspond to the ASQ age intervals. Training on the ASQ is also provided through the Michigan Public Health Institute. A videotape is available that provides guidance on using the ASQ system in a home visiting context.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Not described.

References:

Bricker, Diane, J. Squires, R. Kaminski, and L. Mounts. "The Validity, Reliability, and Cost of a Parent-Completed Questionnaire System to Evaluate At-Risk Infants." *Journal of Pediatric Psychology*, vol. 13, no.5, 1988, pp.5-68.

Squires, Jane, D. Bricker, and L. Potter. *Ages & Stages Questionnaires (ASQ): A Parent-Completed, Child-Monitoring System*. Baltimore, MD: Paul H. Brookes Publishing Co., 1999.

AGES AND STAGES QUESTIONNAIRES: SOCIAL-EMOTIONAL (ASQ:SE) A PARENT-COMPLETED, CHILD-MONITORING SYSTEM FOR SOCIAL-EMOTIONAL BEHAVIORS, 2002

Authors:

Jane Squires, Diane Bricker, and Elizabeth Twombly

Publisher:

Paul H. Brookes Publishing Co.

1-800-638-3775

www.brookespublishing.com

Initial Material Cost:

Questionnaires and User's Guide: \$125

Representativeness of Norming Sample: Compared with 2000 Census figures, the normative sample underrepresents Caucasians and overrepresents individuals of mixed ethnicity and has higher percentage of well-educated mothers and low-income families.

Languages:

English and Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval: 3 to 66 months; administered within 3 months of the target ages of 6, 12, 18, 24, and 30 months, and within 6 months of the target ages of 36, 48, and 60 months

Personnel, Training, Administration, and Scoring Requirements:

The ASQ:SE can be administered by parents, child care providers, and preschool teachers (10 to 15 minutes per questionnaire). Ideally, program staff will train parents on administering the ASQ:SE; training takes approximately 2 to 3 hours. Scoring should be done by a paraprofessional, and should only take a few minutes per questionnaire.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.50 or higher)

Norming Sample Characteristics: 2 (not nationally representative, normed within the past 15 years)

Ease of Administration and Scoring: 3 (self-administered by parent but scoring by paraprofessional recommended)

Description: This series of eight parent-completed questionnaires with 22 to 36 items in each questionnaire helps determine children's progress in their social-emotional behavior. Each questionnaire can be used within 3 months of the target age (for the 6- through 30-month questionnaires) or 6 months of the target age (for the

36- through 60-month questionnaires). The questionnaires focus on seven behavioral areas: (1) self-regulation, (2) compliance, (3) communication, (4) adaptive functioning, (5) autonomy, (6) affect, and (7) interaction with people. Each questionnaire is written at a 5th- to 6th-grade reading level. The ASQ:SE can be used to screen

for social-emotional development problems at one point in time or to monitor a child repeatedly at different intervals. The publisher recommends that the ASQ:SE be used in conjunction with a developmental screening tool that provides information on the child's communication, motor, and cognitive functioning.

Uses of Information: The ASQ:SE was developed to complement the ASQ by providing information specifically addressing the social and emotional behavior of children ranging in age from 3 to 66 months. The ASQ:SE is a screening tool that helps practitioners identify infants and young children whose social or emotional development requires further evaluation to determine whether referral for intervention services is necessary.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): the alphas for the questionnaires were .82 overall, .69 (6-month), .67 (12-month), 18-month (.81); 24-month (.80); 30-month (.88); 36-month (.89); 48-month (.91); 60-month (.91). (2) Test-retest reliability, with one to three weeks between tests: percent agreement between scores by the same rater on two occasions is 94 percent. (3) Inter-rater reliability: no information available.

Validity: (1) Concurrent validity: percent agreement of ASQ:SE with similar established tools ranged from 81 to 95 percent and was 93 percent overall. (2) Predictive validity: no information available.

Method of Scoring: Scoring can be done by paraprofessional or professional staff. Scoring options for the items in the ASQ:SE are "most of

the time," "sometimes," and "rarely or never." Each response is converted to a numerical value. The numerical values are totaled and compared with the empirically derived cutoff score (for that particular questionnaire interval) that indicates whether a child should receive further in-depth evaluation. The reproducible scoring sheets all include referral considerations that help determine whether the child needs further evaluation.

Interpretability: An Information Summary Sheet is provided to assist program staff with scoring and summarizing assessment information and providing them with a summary of the child's performance on the questionnaire. The Information Summary Sheet contains instructions for scoring the questionnaire, a chart indicating cutoff scores for referrals, and a list of considerations prior to making referrals to mental health professionals. Children whose scores are at or greater than the cutoff point should be considered for further evaluation or referral, and children with scores below the cutoff point can be monitored with another ASQ:SE in 6 to 12 months.

Training Support: The User's Guide contains complete instructions for training on the ASQ:SE, setting up the assessment, and conducting it. It provides instructions for administering the questionnaires with sensitivity to children's environmental, cultural, and social-emotional differences. Other support materials include compilation of detailed technical data on how the system was developed and tested, case examples, and creative activities and lists of social-emotional behaviors professionals can share with parents for use with

each age group. The publisher also offers customized training seminars to provide guidance on using this assessment tool.

Adaptations/Special Instructions for Individuals with Disabilities: The ASQ:SE User's Guide briefly mentions the importance of interpreting assessment information within the context of the specific child's health, development, and family/cultural factors. The guide also describes factors to consider before making a referral based on the ASQ:SE assessment.

Report Preparation Support: None described.

References:

Squires, Jane, Diane Bricker, and Elizabeth Twombly. *Ages And Stages Questionnaires: Social-Emotional (ASQ:SE), A Parent-Completed, Child-Monitoring System For Social-Emotional Behaviors*. Baltimore: Paul H. Brookes Publishing Co., 2002.

Squires, Jane, Diane Bricker, and Elizabeth Twombly. *Ages And Stages Questionnaires: Social-Emotional (ASQ:SE), A Parent-Completed, Child-Monitoring System For Social-Emotional Behaviors, User's Guide*. Baltimore: Paul H. Brookes Publishing Co., 2002.

For information on continuing research and adjustments in cutoff points on the ASQ:SE, refer to www.brookespublishing.com/asqse.

ASSESSMENT, EVALUATION, AND PROGRAMMING SYSTEM (AEPS) MEASUREMENT FOR BIRTH TO THREE YEARS, 1993

Authors:

Diane Bricker, Juliann Cripe, Kristine Slentz

Publisher:

Paul H. Brookes Publishing Co.
(800) 638-3775
www.brookespublishing.com

Initial Material Cost:

AEPS™ Birth to Three set (Administration Guide, Test, Curriculum for Birth to Three Years): \$150

Representativeness of Norming Sample: No norming sample.

Languages:

English¹

Type of Assessment:

Observation, direct assessment (to elicit a behavior), and parent, caregiver, or therapist report

Age Range and Administration Interval: Children whose developmental age is 3 years or less. May be used for children whose chronological age is 6 years or less. Readministered at 3- or 4-month intervals.

Personnel, Training, Administration, and Scoring Requirements:

The AEPS can be used by both direct service personnel and specialists. Administration time may range from 1 to 2 hours for the initial assessment and 15 to 30 minutes for subsequent assessments, depending on the child's level of functioning and the user's familiarity with the AESP and the child.. Quarterly or yearly followups generally take half the time of the initial assessment.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher).

Validity: 3 (.5 or higher for content validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic clerical skills)

Description: The Assessment, Evaluation, and Programming System (AEPS) Measurement for Birth to Three Years is a criterion-referenced assessment tool that is designed to help early

interventionists improve their assessments of the abilities and needs of young children who have disabilities or are at risk for developmental delays. The AEPS test was designed to be used in con-

¹ The first edition also has Spanish translations of the Family Interest Survey and the Family Report.

junction with the AEPS Curriculum for Birth to Three Years or other similar curricula. It covers the developmental progress of children's functional skills in six key domains (fine motor, gross motor, adaptive, cognitive, social-communication, and social development). Each domain is divided into strands, which consist of related groups of behavior divided into common categories. Each strand has a series of goals and discrete objectives that lead up to the goal. The strands, goals, and objectives are developmentally sequenced. Objectives and goals are either observed, elicited, or recorded based on parent, caregiver, or therapist report.

The AEPS encourages family participation in the assessment through the use of family-focused materials, such as the family report, planning guide, child progress record, and family interest survey.

Uses of Information: The AEPS is an assessment/evaluation tool that is used to create individual evaluation programs and intervention plans, known as Individualized Educational Program/Individualized Family Service Plans (IEP/IFSP). The test is used to provide a baseline on the child's functioning to help develop the intervention curriculum and to measure the child's developmental progress over time.

Reliability: Several research samples were drawn from children ages 2 months to 6 years in early intervention programs in Oregon, Washington, Idaho, and Arkansas, and 48 interventionists in Vermont, Iowa, Oregon, and British Columbia. (1) Inter-rater reliability: Pearson

product moment correlations for individual domains ranged from .71 for the Social Domain to .96 for the Gross Motor Domain. Mean correlation for all domains was .88. Total test score correlation was .97. (2) Test-retest reliability (1-2 week interval): Pearson Product Moment correlation for domains ranged from .77 for the Social Domain to .95 for the Gross Motor Domain, with a correlation of .88 for all domains. Total test agreement was .95.

Validity: (1) Congruent Validity: Pearson Product Moment correlations with the Bayley Scales of Infant Development Mental Age and Motor Age were .93 and .88, respectively. Correlation with the Gesell Developmental Scale Maturity Age scores was .51.

Method of Scoring: Each of the six domains has a specific recording form. Items are marked as "pass consistently" (2), "inconsistent performance" (1), and "does not pass" (0). Specific criteria are provided for each goal and objective. In addition to scoring each of the items, a qualifying note is attached to each item goal and objective. Items are marked as "assistance provided" (A), "behavior interfered" (B), "reported assessment" (R), "modification/adaptation" (M), and "direct test" (D).

Scoring can be done two ways: (1) a total score is computed for domains by counting the number of goals and objectives scored with a "2" in each domain. For the total frequency, the domain scores are added together. The number of "1" scores are computed in the same way. (2) The percentage of items scored with a "2" or "1" can

also be calculated by dividing the total “2” scores by the total number of items in the domain and the total “1” scores by the total number of items in the domain. For an overall percent score, the total number of items scored with a “2” across the domains is multiplied by 2 and divided by 456 (total number possible).

The AEPS Family Interest Survey and AEPS Family Report, which are family-centered materials that can be used in conjunction with the AEPS Test, are completed by families and have separate scoring guidelines.

Interpretability: No instructions provided. However, the scoring of the instrument will inform the interventionist how well the child is performing the observed skill. The test results expand the quantity and quality of developmental information and help professionals develop IFSP or IEP goals and objectives.

Training Support: “Brookes on Location” professional development seminar, AEPS™: A Linked System of Assessment, Intervention, and Evaluation, is available through the publisher (www.brookespublishing.com).

Adaptations/Special Instructions for Individuals with Disabilities: The AEPS was designed for use with populations of children who are at risk and who have disabilities. For children who have severe impairments, general modifications are required. For children with severe disabilities, the AEPS test objectives should be used more as goals, and the associated curricular programming steps as objectives. General adaptation guidelines are provided for children with visual, hearing, and motor impairments.

Report Preparation Support: Data Recording Forms are provided for scoring tests, graphing results, tracking scores, and recording comments. A Child Progress Record is available to track progress on strand objectives.

References:

Cripe, Juliann, Kristine Slentz, and Diane Bricker. AEPS Curriculum for Birth to Three Years, Volume 2. Baltimore, MD: Paul H. Brookes Publishing Co., Inc., 1993.

BATTELLE DEVELOPMENTAL INVENTORY (BDI), 1984

Authors:

J. Newborg, J.R. Stock, & J. Wnek (initial development); J. Guidubaldi (pilot norming study); J.S. Svinicki (completion and standardization)

Publisher:

Riverside Publishing Co.
800-323-9540
www.riverpub.com

Initial Material Cost:

Examiner's manual: \$58
Scoring booklets (15): \$35
Examiner's manual for screening test: \$56
Screening test booklets (30): \$41
Overview videotape: \$58

Representativeness of Norming Sample: National norming sample of 800 children from birth to 8 years of age, stratified according to geographical region, age, race, and gender; 75 percent urban and 25 percent rural; 28 test sites in 24 states. Distribution of sample closely represents the four major geographical regions of the United States.

Languages:

English

Type of Assessment:

Direct Child Assessment, Observation, and Parent Interview

Age Range and Administration Interval:

Birth to 8 years.

Personnel, Training, Administration, and Scoring Requirements:

Can be administered by paraprofessionals ("nonpsychologists") and is intended for use by infant, preschool, primary, and special education teachers. Important that examiners have supervised practice in administering BDI for children with disabilities across age span.

BDI Screening Test takes 10 to 15 minutes for children under 3 and over 5 years of age and 20 to 30 minutes for children between the ages of 3 and 5. The full BDI can be administered in about 1 hour for children under 3 and over 5 years of age and in 1.5 to 2 hours for children between 3 and 5 years.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (test-retest, .65 or higher)

Validity: 3 (.5 or higher for concurrent)

Norming Sample Characteristics: 2 (older than 15 years, nationally representative)

Ease of Administration and Scoring: 2 (administered and scored by paraprofessionals)

Description: The BDI assesses children from birth to 8 years of age on the following five domains: Personal-Social, Adaptive, Motor, Communication, and Cognitive. The BDI Screening test contains 96 items and represents a

subset of the full battery, which is comprised of 341 items. Within each domain, the items are assigned age levels and organized sequentially into subdomains. The BDI is norm-referenced and helps to identify young children with special

needs and assess the functional abilities of these children, as well as children without special needs. Child diagnostic information for the full BDI is presented in the form of age equivalents, percentiles, and standard scores (that is, developmental quotients, z-scores, T-scores, and normal curve equivalents) for each of the major domains and subdomains. This information is available for the Screening Test as age equivalents and cutoff scores.

Uses of Information: The BDI is primarily used for four purposes: (1) assessment and identification of children with special needs, (2) assessment of school readiness among children without special needs, (3) planning and providing educational instruction, and (4) evaluation of groups of children with special needs.

Reliability: Test-retest (4-month interval) for the total test is .98 for children 0 to 5 and 18 to 23 months old and .99 for children 6 to 11, 12 to 17, 24 to 35, and 36 to 47 months old.

Validity: (1) Content validity: The process for developing the BDI involved identifying skill areas to be assessed, selecting or developing the test items, and verifying the content validity of the results with review by content experts. (2) Construct validity: Factor analysis and the intercorrelations between the domains and subdomains supported the factorial validity and conceptual structure of the BDI. (3) Concurrent validity: Measures on the BDI relate well to other instruments, including the Vineland Social Maturity Scale (Doll 1965), and the Developmental Activities Screening Inventory (DASI; Dubose & Langley 1977), with correlations ranging from .78 to .94. While the BDI is not an intelligence test, it measures motor and language skills and is found to relate moderately well with the Stanford-Binet Intelligence Scale (S-B; Terman & Merrill 1960), with correlations ranging

from .40 to .61. Validity tests were also conducted with the BDI Screening Test. The correlation between the total score on the Screening Test and that of the full battery is .99, indicating that performance on the Screening Test predicts performance on the full BDI.

Method of Scoring: Items are scored on a three-point system, according to whether the child typically completes the item correctly (2), sometimes does so (1), or rarely or never completes the task, even if the child did not have the opportunity to respond (0). The number of allowed trials is presented separately with each item. If a child completes the item correctly on the first trial, he or she receives 2 points and can move on to the next item. Basal rules are established so that test items that are extremely easy for a child need not be administered, while ceiling rules ensure that items that are much too difficult are not administered. A child receives full credit, 2 points per item, for all items that fall below the basal level. The subdomain raw scores for the full BDI battery are obtained by summing the individual item scores from the basal level through the ceiling level and then adding that total to the full credit sum (2 points per item) for items below the basal level. In contrast, for the Screening Test the basal and ceiling rules apply to each domain. Domain raw scores can be obtained by summing the subdomain raw scores, and a total raw score is obtained by summing the five domain raw scores. Once obtained, the raw scores are transferred to the Score Summary and Profile section in the Scoring Booklet.

Interpretability: Tables are used to convert raw scores to percentile rank, age equivalent scores, and the following standard scores: z-scores, T scores, deviation quotients, and normal curve equivalents. The Screening Test cutoff scores for each age group are provided for

three probability levels that correspond to 1.0, 1.5, and 2.0 standard deviations below the mean. In interpreting the full BDI, standard scores are useful for decision making, percentile ranks are useful for reporting information to parents, and age equivalent scores may be required by federal, state, and local policies. Cutoffs are not provided for the full BDI, but the authors recommend following convention and treating standard scores that are 1.5 or more standard deviations below the mean as an indication of a performance deficit. BDI norms should not be used if the BDI has not been administered according to standard procedures. Five case studies on the interpretation of scores are presented in chapter 4 of the manual.

Training Support: An overview videotape is available.

Adaptations/Special Instructions for Individuals with Disabilities: The BDI includes guidelines for assessing children with disabilities so that they are able to respond in a manner that is appropriate for them. Also, most of the items include standardized stimulus/response options for children with visual, hearing, neuromotor, or behavior/emotional needs. Children with special needs are scored according to the same criteria used to score children without disabilities. Adaptations are not made for children who have “no opportunity” for a response on an item due to handicapping conditions, because a score of 0 reflects children’s actual level of functioning.

Report Preparation Support: Guidelines for developing goals and objectives for children with special needs through the Individual Education Plans are found in chapter 5 of the manual.

References:

Doll, E.A. Vineland Social Maturity Scale. Circle Pines, MN: American Guidance Service, 1965.

DuBose, R.F. and M.B. Langley. Developmental Activities Screening Inventory. Hingham, MA: Teaching Resources Corporation, 1977.

Newborg, Jean, John Stock, Linda Wnek, John Guidubaldi, and John Svinicki. Battelle Developmental Inventory with Recalibrated Technical Data and Norms: Examiner’s Manual. Riverside Publishing, 1984.

Newborg, Jean, John Stock, Linda Wnek, John Guidubaldi, and John Svinicki. Battelle Developmental Inventory with Calibrated Technical Data and Norms: Screening Test. Riverside Publishing, 1984.

Terman, L.M. and M.A. Merrill. Stanford-Binet Intelligence Scale, Third Revision (Form L-M). Boston, MA: Houghton Mifflin, 1960.

BAYLEY SCALE FOR INFANT DEVELOPMENT, SECOND EDITION (BSID-II), 1993

Authors:

Nancy Bayley

Publisher:

The Psychological Corporation

(800) 872-1726

www.psychcorp.com

Initial Material Cost:

Complete Kit: \$950 (includes manual, stimulus booklet, 25 mental scale record forms, 25 motor scale record forms, and 25 behavior rating scale record forms, visual stimulus cards, map, and all necessary manipulatives, in a soft-side carrying case)

Representativeness of Norming Sample:

The norming sample was a national, stratified random sample of 1,700 children ages 1 to 42 months. The stratifying variables were age, sex, region, race/ethnicity, and parent education. The sample consisted of 17 age groups ranging from one month among the younger children and increasing to three months among the older children, each with 100 children.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

1 to 42 months

Personnel, Training, Administration, and Scoring Requirements:

Examiner should be trained and experienced in administering and interpreting comprehensive developmental assessments. BSID-II takes 15 to 35 minutes to administer to children under 15 months and up to 60 minutes to children older than 15 months.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher)

Validity: 2 (Mental Scale .5 or higher, Motor Scale <.5 (concurrent))

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally representative sample)

Ease of Administration and Scoring: 3 (requires a highly trained individual).

Description: BSID-II is an individually administered examination that assesses the developmental functioning of infants and children ages 1 to 42 months. BSID-II presents infants with situations and tasks designed to produce an observable set of behavioral responses. The observed responses are scored on complementary

development scales—mental scale, motor scale, and behavior rating scale (BRS). The mental scale assesses the child's level of cognitive, language, and personal-social development. The motor scale assesses the child's level of fine and gross motor development. The BRS assesses the child's behavior during the testing situation,

which facilitates interpretation of the mental and motor scales. The Bayley Infant Neurodevelopmental Screener, which contains 11 to 13 items selected from BSID-II, allows programs with high caseloads to screen infants 3 to 24 months for neurological impairment or developmental delay in 10 to 20 minutes.

Uses of Information: The BSID-II was designed for use in identifying areas of relative impairment or delay, developing curricula for

tests of school-age abilities, the BSID-II is not intended to serve as an intelligence test.

Reliability: (1) Internal consistency (Cronbach's alpha): averages across all age groups were .88 for the mental scale, .84 for the motor scale, and .88 for the BRS total score. (2) Test-retest reliability, with a median 4-day interval between tests for children ages 1 and 12 months: .83 for the mental scale, .77 for the motor scale, and .55 for BRS total score at 1 month and .90 at

	Mental Scale	Motor Scale	Behavior Rating Scale
Internal Consistency (Average Cronbach's Alpha Across Age Groups)	.88	.84	.88
Internal Consistency (Average Cronbach's Alpha Across Age Groups)	.87	.78	
1 to 12 months of age	.83	.77	.55 to .90
24 to 42 months of age	.91	.79	.60
Inter-Rater Reliability	.96	.75	.47 to 1.00

interventions, and assessing the outcome of such interventions. The scales should not be used to measure a child's deficit in a specific skill area or to obtain a norm-referenced score for a child with severe sensory or physical impairments. Also, although items on the mental and motor scales for older children are similar to items found on

12 months of age; for children ages 24 and 42 months: .91 for the mental scale, .79 for the motor scale, and .60 for the BRS total score. Overall, the test-retest reliability coefficients were .87 on the mental scale and .78 on the motor scale. (3) The inter-rater reliability for the mental scale was .96 and for the motor scale, .75. Inter-

rater reliability coefficients for the IBR (now the BRS) ranged from .47 to 1.00.

Validity: (1) Concurrent validity: the correlation between BSID-II and BSID was .62 on the Mental Development Index (MDI) and .63 on the Psychomotor Development Index (PDI). The BSID-II MDI scores were found to be highly correlated and the PDI scores moderately correlated with other scales of development, language, and intelligence. The MDI was found to have a correlation of .79 with the McCarthy Scales of Children Abilities (MSCA) general cognitive index and correlations of .73, .73, and .63 with the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) full scale IQ, Verbal IQ, and Performance IQ, respectively. The PDI was found to have a correlation of .45 with the MSCA's general cognitive index and .41, .39, and .37 with the WPPSI-R full scale IQ, Verbal IQ, and Performance IQ, respectively. Tests found smaller correlation coefficients between BSID-II and the Differential Ability Scales and the Preschool Language Scale-3. Once again, the strength of the relationships was stronger with the MDI. Finally, tests found the BSID-II and the Denver Developmental Screening Test-II to be in agreement in classifying children approximately 80 percent of the time.

Method of Scoring: For each item, the manual provides scoring instructions. The examiner scores an item by entering one of a number of scoring options. By converting mental and motor raw scores into MDI and PDI scores, the examiner can compare a child's performance to the per-

formance of children of similar ages. Another table provides the age-appropriate percentile ranking for the child's BRS score.

Interpretability: BSID-II provides instruction on how to interpret the assessment results and provides three case studies as examples. Additional information on interpreting BSID-II scores is provided in the book, *Essentials of Bayley Scales of Infant Development II Assessment*.

Training Support: The manual contains a chapter that provides instructions on administering and scoring the BSID-II assessments. In addition, each item in the scale has directions for administering and scoring the item.

Adaptations/Special Instructions for Individuals with Disabilities: The manual includes a section on the administration of the scales to children with one or more physical or perceptual impairments.

Report Preparation Support: The manual provides case studies that serve as examples of how information from various sources may be integrated, interpreted, and presented.

References:

Bayley, Nancy. *Bayley Scales of Infant Development, Second Edition*. San Antonio, TX: The Psychological Corporation, 1993.

Black, Maureen M. and Kathleen Matula. *Essentials of Bayley Scales of Infant Development II Assessment*. San Antonio, TX: The Psychological Corporation, 1999.

BRIGANCE DIAGNOSTIC INVENTORY OF EARLY DEVELOPMENT, REVISED EDITION (BDIED-R), 1991

Authors:

Albert H. Brigance

Publisher:

Curriculum Associates

800-225-0248

www.curricassoc.com

Initial Material Cost:

Inventory of Early Development-Revised:

\$124

(Optional) IED-Revised Testing Accessories Kit: \$60

IED-Revised Developmental Record Book: \$27

IED-Revised Developmental Record Book 100-Pack:
\$249

Representativeness of Norming Sample:

None; strictly criterion-based.

Languages:

English

Type of Assessment:

Direct child and parent assessment; parent observations

Age Range and Administration Interval:

Birth to 7 developmental years of age.

Personnel, Training, Administration, and Scoring Requirements:

Examiner should have knowledge of child development and be familiar with the procedures in the manual.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher).

Validity: 2 (concurrent and predictive validity not described; content validity >.90)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a paraprofessional with professional supervision).

Description: The BDIED-R assesses children whose developmental age is between birth and 7 years on their performance across over 200 skills within the following 11 developmental domains: preambulatory motor, gross motor, fine motor, self-help, speech and language, general knowledge and comprehension, social and emotional development, readiness, basic reading skills, manuscript writing, and basic math. Administration of the BDIED-R can be adapted and used to accommodate different settings. The examiner can select the skill domains to be assessed, the skill

sequences to administer, and the method for administering the assessment (direct child assessment, caregiver observation, parent interview). The assessment also uses materials that can typically be found in settings serving infants and children; however, an optional Testing Accessories Kit can be purchased from the publisher if the examiner does not have the materials.

Uses of Information: The BDIED-R is used for four main purposes: (1) assessment, (2) diagnosis, (3) record-keeping, and (4) instructional planning.

Reliability: The BDIED-R is criterion-referenced, and reliability tests can be found in the references that were used to develop the assessment. (1) Internal consistency (Cronbach's alpha) across all ages ranges from .94 to .99. Specifically, it is .99 for section A, .96 for section B, .98 for section C, .94 for section D, .97 for section E, .98 for section F, .99 for section G, .99 for section H, .99 for section I, .98 for section J, and .95 for section K. (2) Test-retest reliability across all ages ranges from .77 to .99. It is .77 for section A, .90 for section B, .97 for section C, .98 for section D, .99 for section E, .81 for section F, .99 for section G, .97 for section H, .94 for section I, .91 for section J, and .87 for section K.

Validity: Content validity is based on a review of relevant literature, and responses and critique to the original BDIED. Hambleton's model of item-objective congruence was used to determine content validity, and all of the criterion-referenced items included in the instrument received a score of .90 or higher. The manual provides references that were used to validate the skill sequences and the developmental age.

Method of Scoring: A child receives credit for a skill by mastering the skill. The manual provides the criteria for demonstrating mastery in each skill sequence. The examiner records the child's progress and develops an instructional plan and set of objectives accordingly.

Interpretability: The manual provides no instructions on how to interpret the results, except to caution against the rigid adherence to the developmental age associated with the skills and to recommend tracking the progress in mastering skills.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: There are no specific adaptations made for children with special needs. However, the BDIED-R can be adapted for use with these children, because it accommodates various response styles. It is more effectively used with children who have mild to moderate, rather than severe, disabilities.

Report Preparation Support: Computer-based programs are available so that child assessment data can be translated directly into BDIED-R plans, which are used to develop goals, objectives, and Individualized Education Programs.

References:

Bagnato, S.J., J.T. Neisworth, and S.M. Muson. *LINKing Assessment and Early Intervention: An Authentic Curriculum-Based Approach*. Baltimore, MD: Brookes, 1997.

CAREY TEMPERAMENT SCALES (CTS), 2000

Authors:

William B. Carey

Publisher:

Behavioral-Developmental Initiatives

(800) 405-2313

www.b-di.com

Initial Material Cost:

Specimen set: \$60 (includes a sample of each of the five CTS questionnaires, with scoring and profile sheets, and the Test Manual)

Representativeness of Norming Sample:

Not nationally representative; it was normed on primarily a white middle class population living in the eastern United States.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

1 month to 12 years

Questionnaires are available for the following age ranges: 1 to 4 months, 4 to 11 months, 1 to 2 years, 3 to 7 years, 8 to 12 years

Personnel, Training, Administration, and Scoring Requirements:

Intended for professional use by persons licensed or certified to provide care to children and their parents; administration time is 20 minutes and requires an early high school reading level; scoring time is 15 to 20 minutes for hand scoring and 4 minutes for computer scoring; a professional is needed for scoring and interpretation.

Summary

Initial Material Cost: 1 (< \$100)

Reliability: 2 (all or mostly under .65)

Validity: 1 (none described)

Norming Sample Characteristics: 2 (not nationally representative, normed within the past 15 years)

Ease of Administration and Scoring: 2 (self-administered, but scored by a professional)

Description: The Carey Temperament Scales (CTS) are sets of items for obtaining parent report of a child's temperament. It can be used in research and/or clinical practice. The CTS uses 75 to 100 descriptions of behavior to assess the 9 New York Longitudinal Study characteristics of temperament: (1) activity level, (2) rhythmicity,

(3) approach-withdrawal, (4) adaptability, (5) intensity, (6) mood, (7) attention span and persistence, (8) distractibility, and (9) sensory threshold. CTS is comprised of 5 different questionnaires, three of which are particularly relevant to Early Head Start programs. These are the Early Infant Temperament Questionnaire (EITQ) for

infants ages 1 to 4 months, the Revised Infant Temperament Questionnaire (RITQ) for infants ages 4 to 8 months (and applicable, but not normed, for ages 9 to 11 months), and the Toddler Temperament Scale (TTS) for children ages 1 to 2 years. The other two scales, the Behavioral Style Questionnaire (BSQ) and Middle Childhood Temperament Questionnaire (MCTQ) are for children ages 3 to 12. Each questionnaire contains up to 100 items that are rated on a 6-point scale of frequency ranging from almost never to almost always. These instruments are designed for caregivers who spend a substantial amount of time with the child being assessed.

Uses of Information: The CTS can help caregivers understand a child's temperament and behavioral style. The scales can also be used by caregivers to help place a child in an environment more suitable to the child's temperament or to adapt the environment (including the home and parenting strategies) to the child's temperament. Temperament itself is not considered amenable to intervention.

Reliability: (1) Internal consistency (Cronbach's alphas): EITQ: scale ranged from .43 to .76 (median = .62); RITQ: scale ranged from .49 to .71 (median = .57); TTS: scale ranged from .53 to .86 (median = .70); BSQ scale ranged from .47 to .80 (median = .70); MCTQ scale ranged from .71 to .83 (median = .82). (2) Test-retest reliability: EITQ (20 day test interval): scale ranged from .64 to .79 (median = .68); RITQ (25 day interval): scale ranged from .66 to .81 (median = .75); TTS (1 month interval): scale ranged

from .69 to .89 (median = .81); BSQ (1 month interval): scale ranged from .67 to .94 (scale median = .81); MCTQ (75 day interval): scale ranged from .79 to .93 (median = .88).

Validity: Literature on the clinical evidence for validity and appropriate use of temperament data in practice can be found in *Coping with Children's Temperament* (1995), written by Carey and McDevitt or in *Developmental-Behavioral Pediatrics* (1992), edited by Levine, Carey, and Crocker.

Method of Scoring: The CTS can be hand or computer scored. Items are tabulated to yield a category score for each of the nine areas, which are then compared to the norms for the category. If using the software, the Professional Report includes the temperament profile, raw and standardized scores, individualized interpretive report and validity checks for social desirability, missing data and ratings/perceptions discrepancies. The Caregiver Report contains the temperament profile and an interpretive report of scores written for the caregiver and personalized with the child's name and gender.

Interpretability: Category scores for each of the nine areas can be compared to norms for the category. The manual gives instructions for interpreting the results, depending on whether or not the computer or manual scoring is used. In addition, the authors stress the importance of supplementing the results from the CTS with information gathered from interviews, observations, and other information collected by trained professionals.

Training Support: CTS practice sets are available through the publisher. Individuals with questions may email publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described. However, a younger age questionnaire can be used for individuals with mild delays.

Report Preparation Support: The manual states that the written report or profile should not be automatically given to caregivers because they may not have sufficient information about the limitations of the information. The professional should exercise his or her judgment when deciding whether or not to share the written report or profile.

References:

Carey, William B. *The Carey Temperament Scales Test Manual*. Scottsdale, AZ: Behavioral-Developmental Initiatives, 2000.

Carey, W.B. and McDevitt, S.C. *Coping With Children's Temperament: A Guide for Professionals*. New York: Basic Books, 1995.

Levine, M.D., W.B. Carey, and A.C. Crocker. *Developmental-Behavioral Pediatrics*, 2nd ed. Philadelphia, PA: W.B. Saunders, 1992.

THE CAROLINA CURRICULUM FOR INFANTS AND TODDLERS WITH SPECIAL NEEDS (CCITSN), ASSESSMENT LOG, SECOND EDITION, 1991

Authors:

Nancy Johnson-Martin, Kenneth Jens, Susan Attermeier, and Bonnie Hacker.

Publisher:

Paul H. Brookes Publishing Co.
www.brookespublishing.com
1-800-638-3775

Initial Material Cost:

\$43

Representativeness of Norming Sample: None
(criterion-referenced)

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 2 developmental years with mild to severe special needs

Personnel, Training, Administration, and Scoring Requirements:

Can be used by both professionals and paraprofessionals. Effort made to avoid use of technical jargon in materials to encourage broader usage. Requires informal observation period in which the examiner can assess and score child in about 15 to 20 minutes. The Assessment Log is designed to be used in conjunction with the Carolina Curriculum.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 1 (content validity only described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a professional or paraprofessional)

Description: The Carolina Curriculum for Infant and Toddlers with Special Needs (CCITSN) is designed for use with infants from birth to 2 years developmental age who have mild to severe special needs. The curriculum covers 6 developmental domains (cognition, communication, social/adaptation, fine motor, and gross motor) that are divided into 26 teaching areas (or sequences). CCITSN has an Assessment Log that enumerates 26 sequences and the specific skills

under each sequence are ordered sequentially in terms of the expected development of children. The number of items a child is assessed on is at the discretion of a professional or paraprofessional, who administers the CCITSN Assessment Log through an informal observation of the parent-child interaction. This format is preferred to a more clinical, structured approach to assessment.

Uses of Information: The CCITSN Assessment Log is used to identify the curriculum

entry point, to inform the intervention plan, and to monitor progress in accomplishing the skills covered by the curriculum.

Reliability: None described.

Validity: Content validity: the selection of items for inclusion in the curriculum Assessment Log was accomplished through a review of norm-referenced tests of development. A multi-disciplinary panel of specialists helped in the final selection process. The curriculum was field tested in 22 intervention programs in North Carolina and in 10 national sites, and the interventionists found it to be useful both for assessing infants with disabilities and for developing their intervention programs.

Method of Scoring: Behaviors on the question items are scored as either typical of the child (+), emergent (+-), or never observed (-). Once behaviors have been scored, the Developmental Progress Charts are used to chart assessment results and develop a profile of the child's skills. There is a blank box on the Developmental Progress Chart to correspond to each item on the Assessment Log, which is colored in completely when an item is passed. If the skill is inconsistently performed or emerging, the space is partially colored in.

Interpretability: Generally, a child's performance can be assessed based on his or her performance of the three items after the first failure

and the three items before the first success in each sequence. The Assessment Log then serves as a basis for intervention using the Curriculum Sequences by selecting the first activity the child failed or that was just emerging in each sequence.

Training Support: "Brookes on Location" professional development seminar, Using The Carolina Curriculum to Assess and Intervene with Young Children with Special Needs, is available through the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: The CCITSN is designed specifically to optimize the development of children with mild to severe special needs. Interventions, and their associated assessment, are tailored to the child's impairment, and the standard approach will be modified if a handicapping condition makes it inappropriate. The curriculum has special needs options and adaptations available for those with vision, motor or hearing needs.

Report Preparation Support: A profile of the child's skills can be obtained by completing the Developmental Progress Chart.

References:

Johnson-Martin, Nancy, Kenneth Jens, Susan Attermeier, and Bonnie Hacker. 2001. The Carolina Curriculum for Infants and Toddlers with Special Needs, Second Edition. Baltimore, MD: Paul H. Brookes Publishing Co.

DENVER II DEVELOPMENT SCREENING TEST (DDST-II), 1989

Authors:

William K. Frankenburg and J.B. Dodds

Publisher:

Denver Developmental Materials, Inc.
1-800-419-4729 • 303-355-4729 • www.denverii.com

Initial Material Cost:

Denver II Test Kit (includes 100 forms, Screening Manual, and test items): \$84 Training Video: \$215 (purchase) or \$90/week (on-site rental)

Representativeness of Norming Sample:

The English version of the test was normed from 1987 to 1989 on a quota sample of 2,096 English-speaking children in Colorado with no obvious special needs. These children were of varying ages (between 2 weeks and 6.5 years), levels of maternal education, places of residence and cultural backgrounds. The Denver norming sample is representative of Colorado children (from 1980 US Census), and slightly overrepresents Hispanic infants, and underrepresents African American infants. However, when comparing the Colorado average 90% norms with the theoretical US composite norms, there were no clinically significant differences. The DDST-II Spanish version was not normed on Spanish-speaking children, but is a direct translation.

Languages:

English and Spanish (translation of directions and test forms)

Type of Assessment:

Direct child assessment and parent report

Age Range and Administration Interval:

Birth to 6 years

Personnel, Training, Administration, and Scoring Requirements:

Administration of the DDST-II and scoring of the DDST-II and PDQ-II require a professional or paraprofessional. The manual suggests that users carefully review the manual, review the training videotape, and practice testing children of various age groups in order to properly administer and interpret the DDST-II. A two-day training is also suggested. The test takes 10 to 20 minutes to administer, and 1 to 2 minutes to score. The Prescreening Developmental Questionnaire (PDQ-II) takes about 10 to 15 minutes to complete.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (both inter-rater and test-retest reliability have high percent agreement)

Validity: 1 (none described)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (scoring the DDST-II and the PDQ-II requires a highly trained individual)

Description: The DDST-II is a 125-item standardized measure that is designed to determine whether a child's development is within the normal range. It includes a set of questions for parents and tests for the child on twenty simple tasks and items that fall into four sectors: Personal-Social (25 items), Fine Motor Adaptive (29 items), Language (39 items), and Gross Motor (32 items). The number of items administered during an assessment will vary with the child's age and ability.

A Prescreening Developmental Questionnaire (PDQ-II) has been developed to help parents quickly identify whether their children need further assessment. The PDQ-II is a pre-screening consisting of 91 parent questions from the DDST-II. It was created for parents to complete easily and quickly to assess whether their children have non-normal scores and need to complete the full DDST-II. The PDQ-II was revised in 1998 and uses the norms developed for the DDST-II. The questionnaires are divided by age range (0 to 9 months, 9 to 24 months, 2 to 4 years, and 4 to 6 years).

Uses of Information: The DDST-II is intended for use as a screening tool to detect developmental delays. The DDST-II provides a clinical impression of a child's overall development and confirms suspected potential developmental difficulties with an objective measure. It can be used to determine how a child compares to other children and identify children for whom additional in-depth assessment should be conducted. The authors do not recommend using it to predict

later development status, as an in-depth assessment of developmental functioning, or to plan individual intervention programs.

Reliability: (1) Internal consistency reliability: no information available. (2) Test-retest reliability: 89 percent agreement between test scores for a 7- to 10-day interval between test administrations by the same tester. (3) Inter-rater reliability: for the standardization sample, percentage agreement between examiners and a criterion observer (inter-rater reliability) ranged from 92 to 98 percent.

Validity: No information available

Method of Scoring: The child's responses are recorded as Pass or Fail on the score sheets. The responses are examined to see if they fall into or outside the normal expected range of success on that item for the child's age (the child is either classified as normal range, suspect, or delayed).

Interpretability: The DDST-II scoring process, which is described in the screening manual, requires that the individual test items be interpreted before the entire test is interpreted. The individual items are classified as: Advanced, Normal, Caution, Delayed, and No Opportunity. The category descriptors for the entire test include: Normal, Abnormal, Questionable, and Untestable.

Training Support: It is suggested that screeners be properly trained and pass the proficiency test before using the DDST-II for clinical purposes. There is a two-day training workshop offered (and outlined in the technical manual).

Adaptations/Special Instructions for Individuals with Disabilities: None mentioned

Report Preparation Support: None mentioned

References:

Frankenburg, William K. and J.B. Dobbs. Denver Developmental Screening Test II- Screening Manual. Denver: Denver Developmental Materials, 1990.

Frankenburg, William K. and J.B. Dobbs.

Denver Developmental Screening Test II- Technical Manual. Denver: Denver Developmental Materials, 1990.

Frankenburg, William K. and J.B. Dobbs. Denver Developmental Screening Test II- Training Videotape. Denver: Denver Developmental Materials, 1993.

Personal correspondence with Beverly Bresnick, DDST-II technical expert and trainer, July 10, 2002.

DEVELOPMENTAL OBSERVATION CHECKLIST SYSTEM (DOCS), 1994

Authors:

Wayne Hresko, Shirley Miguel, Rita Sherbenou, and Steve Burton

Publisher:

Pro-ed

(800) 897-3202

www.proedinc.com/index.html

Initial Material Cost:

Complete DOCS Kit: \$129 (includes Examiner's Manual, 25 Cumulative Profile/Record Forms, 25 Developmental Checklist

Profile/Record Forms, 25 Adjustment Behavior Checklist

Profile/Record Forms, and 25 Parent Stress and Support Checklist Profile/Record Forms)

Representativeness of Norming Sample: DOCS was normed on more than 1,400 children ages birth through 6 years from more than 30 states. Although a random sampling procedure was not used, characteristics of the normative group approximate those for the 1990 U.S. Census data relative to gender, geographic region, race/ethnicity, and urban/rural residence. The tests were conducted between November 1989 and December 1992.

Languages:

English

Type of Assessment:

Parent or caregiver report

Age Range and Administration Interval:

Birth to 6 years

Personnel, Training, Administration, and**Scoring Requirements:**

Examiners should have some training in administering and interpreting assessment instruments. The instrument can be completed by a parent with a fourth grade reading level. It takes 30 minutes to complete and 15 to 20 minutes to score all three checklists.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (about half of the coefficients were <.5; about half were >.5)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (self-administered and scored by a trained individual)

Description: The Developmental Observation Checklist System (DOCS) is a three-part instrument to assess the development of very young children, their ability to adjust to their environment, and the level of stress and support

in their environment. Part I uses the Developmental Checklist (DC), a parent report questionnaire, to assess the child's general development in the areas of cognition, language, social, and motor domains. It is answered in a yes/no

format. Part II uses the Adjustment Behavior Checklist (ABC) to screen for any problematic behaviors in the child's ability to adapt to his/her environment. Part III uses the Parental Stress and Support Checklist (PSSC) to identify family stress regarding the child and support used to mediate the stressors. Both the ABC and the PSSC are scored on a 4-point Likert-type scale.

Uses of Information: The DOCS is used to (1) identify infants and children with developmental delays or deficits in cognitive, language, social, and motor abilities; (2) assess adjustment behavior; (3) determine levels of familial stress and support; (4) facilitate the proper professional referral for the child; (5) serve as a measurement device in research studies, (6) give direction to instructional practice, and (7) document educational progress.

Reliability: (1) Internal consistency reliability (Cronbach alphas): the alphas for age groups between birth and 3 years old were in the mid to high .90s for the DC components, in the .80s for the ABC, and in the low to mid .90s for the PSSC. (2) Test-retest reliability (with a 14- to 21-day interval): coefficients for children ages 2 to 3 ranged from .85 to .91 for the DC component and overall checklists and were .94 and .89 for the ABC and PSSC, respectively. (3) Inter-rater reliability: Parent to caregiver standard score reliability coefficients on the DC component and overall ranged from .91 to .94.

Validity: (1) Concurrent validity: The DC component quotient scores correlation with the Bayley Scale for Infant Development, Expressive

One-Word Picture Vocabulary Test (EOWPVT), Denver Developmental Screening Test-Revised, McCarthy, Receptive-Expressive Emergent Language Test (REEL), Test of Early Language Development-2 (TELD-2), Stanford Binet (SB-4th), Slosson Intelligence Test-Revised (SIT-R), Test of Early Socioemotional Development (TOESD), and Vineland Adaptive Behavior Scale ranged from .35 (Developmental Language Quotient, Developmental Cognition Quotient, and Developmental Cognition Quotient with the McCarthy) to .83 (Developmental Language Quotient with the TELD-2). The correlation of DC component quotient scores with the Parental Stress Inventory (PSI) ranged from -.72 to -.38. The correlations of the ABC and PSSC with the TOESD were .65 and .47, respectively. For the Vineland, the correlations with the ABC and the PSSC were .69 and .51, respectively. The correlations of the ABC and PSSC with the PSI were -.38 and -.72, respectively. The DOCS was also able to differentiate between children with normal development and those with developmental challenges. These validity tests were performed on children between the ages of 3 and 6. (2) Predictive validity: no information available.

Method of Scoring: To score the DC, the examiner needs to find the child's basal and ceiling points. The basal is established when the parent or caregiver marks "Yes" for five items in a row, and the ceiling is established when the parent or caregiver marks "No" for five items in a row. The DC score is the sum of all of the items below the basal (including the five basal items) and the

number of “Yes” responses above the basal and below the ceiling. The ABC and the PSSC have no basals or ceilings. Checkmarks in each column are weighted according to their placement in the scale (responses are assigned a number between 1 and 4) and multiplied by a factor indicated on the response sheet. To compute a raw score for each checklist, the correct responses are summed. Using tables in the manual, the DC component checklist raw scores can be converted into percentiles, standard scores, quotients (a distribution with a mean of 100 and a standard deviation of 15), normal curve equivalents, and age-equivalents (the child’s performance age). The manual also has tables to convert the ABC and PSSC raw scores into percentiles and quotients.

Interpretability: The manual provides guidelines for interpreting DOCS scores, as well as cautions about their limitations. In general, while low DOCS scores may indicate the presence of developmental or environmental issues, they do not provide information on the sources and nature of the issues. The examiner is advised to always consider other sources of information, but especially when the assessment has practical implications for the child.

Training Support: None

Adaptations/Special Instructions for

Individuals with Disabilities: Instructions are given for how to administer the instrument if the individual is blind, illiterate, or below a 4th-grade reading level. In addition, the norms provided are appropriate for normally developing children such as those used in the standardization sample. If an individual’s performance is to be compared with a more specific reference group (for example, deaf, retarded, or children older than 6), the authors state that the suitability of the DOCS for that group should be established before evaluating test performance.

Report Preparation Support: There are instructions in the manual for how to share the results with others, including parents.

References:

Hresko, W.P., S.A. Miguel, R.J. Sherbenou, and S.D. Burton. Developmental Observation Checklist System: A Systems Approach to Assessing Very Young Children Examiner’s Manual. Pro-Ed: Austin, TX, 1994.

DEVELOPMENTAL PROFILE II, 2000

Authors:

Gerald Alpern, Thomas Boll, and Marsha Shearer

Publisher:

Western Psychological Services
(800) 648-8857
www.wpspublish.com

Initial Material Cost:

Developmental Profile II Complete Kit: \$120 (includes Examiner's Manual and 25 Scoring/Profile forms)

Representativeness of Norming Sample:

Not nationally representative and not representative of rural populations. Normed on a sample of normally developed children between birth and 12 _ years from the states of Indiana and Washington and primarily from urban areas. Minorities other than African Americans are under-represented.

Languages:

English

Type of Assessment:

Direct child assessment and/or parent report

Age Range and Administration Interval:

Birth through 7 years for normal children and for handicapped children of any age when their skills are not expected to extend beyond the 9 _ year ceiling.

Personnel, Training, Administration, and Scoring Requirements:

A person with brief training can administer this test in 20 to 40 minutes. Training requires familiarizing oneself with the manual and questionnaires. Scoring takes about 5 minutes. However, the ultimate responsibility for its use and interpretation should be assumed by a professional with clinical training.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (under .5 for concurrent)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic clerical skills)

Description: The Developmental Profile II is a comprehensive assessment of motor, language, personal/self-help, social, and intellectual development for children from birth through 9 _ years. The format is a 186-item inventory designed to assess a child's functional, developmental age level. Three methods of administration are available: (1) a Scoring/Profile Form in which the examiner marks the responses and computes and

profiles the child's scores; (2) on-line administration and scoring; and (3) a mail-in, computer-readable answer sheet. There are 13 age groupings, with approximately 15 items per age group (3 items per scale times 5 scales) each. The test may be administered either in interview format to the parent, as a combination of parent interview and direct testing of the child, or as a self-interview completed by a teacher. The examiner's job

is to establish a baseline and ceiling developmental age for each of five areas: physical age, self-help age, social age, academic age, and communication age.

Uses of Information: The primary uses of the results of the Developmental Profile II are (1) to determine eligibility for receiving special education and/or related services; (2) as a planning tool to develop an individualized educational program consistent with the child's strengths and deficits; (3) to measure the child's progress by comparing profile scores at the beginning of the school year (pretest) with scores achieved at the end of the school year (post-test); and (4) to evaluate an entire educational program or service by comparing the average pretest scores of a group of children with the average posttest scores of the same group.

Reliability: (1) Internal consistency reliability: alpha coefficients for each of the five scales were .79 for physical; .78 for self-help; .82 for social, .87 for academic, and .83 for communication. (2) Test-retest reliability (with intervals of 2 to 3 days): scores were identical for 22 percent of the children; 50 percent of the scores were within 1 item of each other; 68 percent were within 2 items of each other; and 92 percent were within 3 items of each other. (3) Inter-rater reliability: 35 teachers independently scored a mother's response to relevant items. Of the 35 teachers, 25 had identical scores, an additional 5 teachers were within one item, and all 35 were within two items of the model score.

Validity: (1) Concurrent validity: correlation coefficients between the Developmental Profile II and the Stanford-Binet ranged from .45 for the Physical Scale and .76 for the Academic Scale. Correlation coefficients between the Developmental Profile II and the Learning Accomplishment Profile (LAP) ranged from .07 to .48 for the Gross Motor Scale. (2) Predictive validity: no information available. (3) Content validity: literature and existing measures were surveyed to identify developmental skills. In order to develop the items in the Profile, a group of teachers that work with handicapped children used the instrument and provided feedback on the instrument's clarity and usefulness for designing and evaluating instructional interventions. An item analysis was used to insure that items were asked in an age appropriate context. Parents generally provided accurate information about their children's performance. When parental report differed from teacher report or direct assessment of children, parents tended to overestimate their children's skills.

Method of Scoring: The child's general behavior, grade or school placement, and any other pertinent information available should be taken into consideration along with the child's chronological age to determine where in the Profile to begin testing. Items are scored as either "pass" or "fail" depending on whether the child has the skill described. If an item is passed, the number in the Pass column is circled; if the item is failed, the zero in the Fail column is circled. A basal credit is established when all skills at two

consecutive levels are mastered. Then, an additional credit is computed, which is the sum of the months earned over and above the basal credit (the sum of all the numbers in the Pass column higher than the level of the basal credit). The basal credit and additional credit are summed to obtain the scale age in months. By subtracting the chronological age from the scale age, the months differential can be computed. A ceiling developmental age (all items are failed on two consecutive age levels) is also established. If a ceiling level is obtained, an IQ Equivalency Score can also be computed for the child. Additionally, the inventory can be administered and scored on-line.

Interpretability: Complete interpretation of the scales requires comparison of scores to those of the normative sample, assessment of individual strengths and weaknesses, and examination of individual item responses. A child's developmental age in each area can be compared to his or her chronological age. If the child's developmental age is higher than the chronological age, the difference will be positive, indicating that the child may be "advanced" in that skill area. If the opposite is true, the difference will be negative, indicating that the child may be "delayed" in that skill area. A child's developmental age can also be

compared to other children to determine whether observed differences between chronological age and developmental age are important. A child's developmental age can also be used to determine the seriousness of a delay. Case studies are included as examples.

Training Support: None

Adaptations/Special Instructions for Individuals with Disabilities: The instrument may be used with handicapped children of any age when their skills are not expected to extend beyond the 9 1/2 year ceiling. Additionally, there is a case study of a child with suspected hearing loss.

Report Preparation Support:

Communication of test results to parents should focus on the interpretation of the results and their implications rather than reporting specific scores. There is a brief description of a sample Developmental Profile II test report in the manual.

References:

Alpern, Gerald, Thomas Boll, and Marsha Shearer. *Developmental Profile II Manual*. Los Angeles, CA: Western Psychological Services, 2000.

DEVEREUX EARLY CHILDHOOD ASSESSMENT (DECA), 1999

Authors:

Paul LeBuffe and Jack Naglieri

Publisher:

Kaplan Press
(800) 334-2014
www.kaplanco.com

Initial Material Cost:

\$200

Representativeness of Norming Sample:

Two non-randomly selected samples of preschool children ages 2 through 5 from 28 states that closely represent the 1995 U.S. population on such important characteristics as age, gender, geographical region, race/ethnicity, and socioeconomic status. One sample of 2,000 children was used to norm the protective scale and the other sample of 1,108 children was used to norm the behavioral problem scale.

Languages:

English; Spanish (Record Form only)

Type of Assessment:

Observation

Age Range and Administration Interval:

2 to 5 years, interval not prescribed, but there should be at least a four-week interval between assessments by the same adult.

Personnel, Training, Administration, and Scoring Requirements:

DECA raters need to be able to complete all of the items on the assessment and need to have observed the child's behavior for a minimum of two hours per day, two days each week, over a period of four weeks. Raters can include parents or other family members, as well as teachers, and should be able to read at the sixth-grade level. DECA users administer the DECA and interpret the scores. They should be trained to interpret and use standardized assessment instruments, standardized scores, and profiles. They should know how to use and communicate results to family members and service providers. Users often include program directors, lead teachers, preschool mental health or educational consultants, and early childhood special educators.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 2 (internal consistency .65 or higher; test-retest .65 or higher; inter-rater mostly <.65).

Validity: 1 (criterion validity, using contrasted groups approach, is significant, but correlation coefficients not reported.)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative).

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual).

Description: The DECA is a 37-item rating scale designed to evaluate self-protecting factors and behavioral concerns among preschool children ages 2 to 5. A 27-item Total Protective Factors scale that assesses self-protective factors represents a compilation of three subscales: Initiative, Self-control, and Attachment. In addition, a 10-item Behavioral Concern scale assesses challenging and problem behaviors that children may exhibit. Family members or early care and education professionals who have observed the focal child's behavior over a period of at least four weeks can complete the DECA, which yields raw scores, percentile scores, T-scores, normal curve equivalent scores, and individual profiles.

Uses of Information: The DECA is used to (1) develop an individual profile to identify strengths and weakness of self-protective factors in order to develop strategies to strengthen these abilities to reduce behavioral problems of preschool children; (2) develop a classroom profile that identifies the relative strengths of all children in a classroom in order that classroom strategies might be implemented to build upon strengths and promote healthy social and emotional child development; (3) identify children with behavioral problems through the Behavioral Concerns Scale so that interventions can be made before behavioral disorders develop; (4) help Head Start programs to meet Program Performance Standards; (5) assist early childhood programs in building on children's strengths to promote healthy child development; (6) provide programs with an outcome measure that can be compared

over time to evaluate the effectiveness of prevention and intervention strategies; (7) compare scores between different raters for the same child to explore any differences that might exist in a child's behavior across different environments; and (8) provide a measure of self-protective factors in preschool children that can be used for research purposes.

Reliability: (1) Internal consistency (Cronbach's alpha): total protective factors scale alpha is .91 (.76 to .86 for the individual protective scales) for parent raters and .94 (.85 to .90 for the individual scales) for teacher raters and for behavioral concerns, .71 and .80, respectively. (2) test-retest reliability: total protective factors scale reliability coefficient is .74 (.55 to .80 for the individual scales) for parent raters and .94 (.87 to .91 for individual scales) for teacher raters and for behavioral concerns, .55 and .68, respectively. (3) inter-rater reliability: coefficients for total protective factors scale are .21 (.26 to .33 for the individual scales) pairs of teacher raters, .69 (.57 to .77 for the individual scales) among pairs of parent raters, and .29 (.19 to .34 for individual scales) among pairs of parent-teacher raters. Coefficients for behavior concerns are .44 among pairs of parents, .62 among pairs of teachers, and .23 for pairs of parent-teacher raters.

Validity: (1) Content-related validity: comparison to well-established measures or related research findings not possible since DECA is the first behavior rating scale to examine self-protective factors; however, the items selected for the DECA were based on an extensive review of the

literature on resilience, as well as results from focus groups with parents and teachers. Factor analysis procedures were used to select the items for each of the protective scales and the 10 items in the Behavioral Concerns Scale. (2) Criterion validity: The DECA scores of a group of pre-school children with identified emotional and behavioral problems were found to have significantly lower protective factor scale scores and significantly higher behavioral concern than a matched sample of children with no identified emotional and behavioral problems. Using the same samples, the authors correctly predicted group members for 69 percent of the children using the Total Protective Factor Scale and 71 percent of the group membership using the Behavioral Concerns Scale.¹ (3) Construct validity: The authors reported findings consistent with DECA's theoretical construct. Low- to average-risk children with high protective factors had the lowest behavioral concern scores, while high-risk children with low protective factors had the highest behavioral concern scores.

Method of Scoring: Raters complete the Record Form by indicating the number of times they have observed the child performing particular behaviors in the past four months. For each item, they place a checkmark next to one of the following descriptors: never, rarely, occasionally, frequently, or very frequently. The checkmarks

are then transferred onto a separate page of the Record Form, where corresponding boxes have raw score values that correspond to each rating: never = 0, rarely = 1, occasionally = 2, frequently = 3, and very frequently = 4. To score the DECA, examiners copy the raw score value (0-4) from the box with its raw score to an empty box that corresponds to the item being scored. The raw scores for the items that comprise each of the scales are summed to obtain the scale raw scores. The scale raw scores for the Initiative, Self-control, and Attachment scales are then summed to obtain the Total Protective Factors Scale Raw Score, which does not include the Behavioral Concerns scale raw score. The raw scores are converted into T-scores and percentiles through the use of the Individual Profile form or tables in the manual. The manual provides tables to help interpret the differences between scores on the protective subscales, by different raters, and over time.

Interpretability: High scores on the Protective Factor scales indicate that a child is doing well, while high scores on the Behavioral Concerns scale may indicate the need for intervention. It is difficult to interpret the meaning of raw scores, so they are converted into percentile scores, T-scores, or normal curve equivalents. This allows scores to be compared to ratings that children typically receive on the scales. T-scores on the DECA range from 30 to 70, and are classi-

¹Factor analysis is a statistical procedure that examines the strength of the association items have with a shared factor. The researcher assigns the meaning to a factor based on the items that are most strongly correlated to the factor.

fied as Below Average (30-40), Average (41-59), or Above Average (60-70). Above Average scores on the Behavioral Concerns scale and Below Average scores on the Total Protective Factors scale may warrant attention. The DECA scores need to be interpreted in the context of other information, including DECA scores from other individuals, and the cultural and family background.

Training Support: Training support is available through the Devereux Foundation and includes introduction and basic implementation sessions, as well as a train-the-trainer session. The cost of training sessions ranges from \$550 to \$1,100 plus additional fees for travel and preparation. The following technical assistance options are also available, upon request, through the Devereux Foundation: (1) on-site visits with devereux certified trainers who provide one-on-one technical assistance for a cost of \$500 per day plus travel and one day of preparation; (2) video conferences to provide refresher training; (3) tele-

phone conferences with DECA program developers and trainers; and (4) a web-based listserv discussion forum.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Individual and Classroom Profiles are used to graphically display results from the DECA on all five scales. Several examples of DECA Individual Profiles are presented and interpreted, along with suggestions for an intervention plan.

References:

LeBuffe, Paul and Jack Naglieri. Devereux Early Childhood Assessment: User's Guide. Lewisville, NC: Kaplan Press, 1999.

LeBuffe, Paul and Jack Naglieri. Devereux Early Childhood Assessment: Technical Manual. Lewisville, NC: Kaplan Press, 1999.

EARLY COPING INVENTORY (ECI), 1988

Authors:

Shirley Zeitlin, G. Gordon Williamson, and Margery Szczepanski

Publisher:

Scholastic Testing Service, Inc.
www.ststesting.com
1-800-642-6787

Initial Material Cost:

ECI Manual and 20 forms: \$51

Representativeness of Norming Sample: None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

Children whose developmental age is between 4 and 36 months

Personnel, Training, Administration, and Scoring Requirements:

Nonprofessionals who are knowledgeable in infant development may administer and score the inventory. If observers are not familiar with the child, they should observe the child at least 3 times in different situations. Interpretation should be done by a professional with a background in early development and behavioral sciences. Administration time varies depending on the observer's familiarity with the child.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Interrater: 3 (.65 or higher)

Validity: 1 (content validity only reported)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (may be administered by a nonprofessional but scored by a highly trained individual).

Description: The Early Coping Inventory (ECI) is an observation instrument used for assessing the coping-related behavior of children whose chronological or developmental age is between 4 and 36 months. The ECI's 48 items are divided into 3 coping clusters: Sensorimotor Organization, Reactive Behavior, and Self-

Initiated Behavior. Each item is rated on a five-point scale ranging from ineffective coping (1) to consistently effective coping across situations (5).

Uses of Information: Analysis of a child's scores on the instrument provides information about level of coping, style, and specific strengths and weaknesses. The findings can then be used to

create educational and therapeutic interventions. In addition, the ECI can be used to involve parents in its use as a means of increasing knowledge of the child and communication with staff. The ECI can also be used to support staff development and training to increase observation skills, expand their domain of concern, facilitate teamwork, and measure child progress. The manual provides a chapter on how the ECI can be used for these purposes.

Reliability:¹ Reliability was established with a group of observers who completed the ECI after viewing videotapes of four young children twice, with a 6-week interval between viewing. (1) Interrater reliability (using Guildford's formula): At the first viewing, the reliability coefficients for the three coping clusters and the adaptive behavioral index ranged from .80 (sensorimotor organization) to .94 (self-initiative behavior) and at the second viewing from .87 (sensorimotor organization) to .93 (self-initiative behavior). The authors also tested for the level of agreement between the observers' scores with the scores of an expert panel for each of the ECI items. The "concordance index" showed that the agreements in the item scores within each coping cluster ranged from a mean of 41 percent (reactive behavior) to a mean of 52 percent (sensorimotor organization). (2) Test-retest reliability (six-week interval): Friedman's analysis of variance test was used to test for significant differences between ECI test-

retest scores for each child on the coping clusters and the adaptive behavioral index. The authors reported no statistical significant shift in scoring on 11 of the 16 tests.

Validity: (1) Content validity: Item content and definitions of coping constructs were primarily derived from a review of the early childhood, coping-related literature. A panel of six judges then reviewed the selected items for fidelity to the coping constructs. The EIC was then administered to three different samples and the responses were factor analyzed. The results of the factor analysis provided the basis for creating the three coping clusters.

Method of Scoring: Raw score totals are calculated for sensorimotor organization, reactive behavior, and self-initiated behavior by summing the items ratings scale numeric values. A table is used to convert the raw scores into Effectiveness scores, which can be plotted on the Coping Profile and used to compare the child's level of effectiveness in the three categories. Another table converts the sum of the effectiveness scores into an Adaptive Behavior Index score. A list of six to eight of the Most and Least Adaptive Coping Behaviors is also compiled to aid intervention planning.

Interpretability: Higher scores indicate the use of more effective coping behaviors in adapting to stresses in every day living. The Adaptive

¹The reliability tests were conducted using the research version of the ECI, which had 54 items instead of 48 items in the current version.0

Behavior Index indicates the child's general level of effectiveness in using adaptive behaviors to cope and whether or not intervention is needed. The Coping Profile, which graphically displays the effectiveness scores, shows strengths and vulnerabilities in coping behavior. The authors recommend taking advantage of areas of strengths when planning intervention activities to increase competence in the weaker areas. In the same way, the list of the Most and Least Adaptive Coping Behaviors can also be used to develop the intervention plan. The manual provides a table that contains a descriptive interpretation for the Effectiveness and Adaptive Behavior Index numeric scores and several case studies to illustrate how to interpret the results.

Training Support: Two examples are provided to illustrate use of the rating scale and three case studies are given on how to interpret the results. Appendix D of the manual also contains case studies on how to develop intervention plans.

Adaptations/Special Instructions for Individuals with Disabilities: Instructions are given in the manual on how to rate a child with a disability.

Report Preparation Support: None.

References:

Zeitlin, Shirley, G. Gordon Williamson, and Margery Szczepanski. *Early Coping Inventory: A Measure of Adaptive Behavior*. Bensenville, Illinois: Scholastic Testing Service, 1988.

NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT PARENT INTERVIEWS AND CHILD ASSESSMENTS, 1996 - 2001

Authors:

John Love and other project staff, in collaboration with the Early Head Start Research Consortium

Publisher:

Mathematica Policy Research, Inc.
Contact Publications, 609-275-2350,
jallen@mathematica-mpr.com . Lists of measures and copies of the interviews can be found at
www.mathematica-mpr.com or
www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_instruments.html.

Initial Material Cost:

None

Representativeness of Norming Sample:

None described

Languages:

English, Spanish

Type of Assessment:

Parent report, observation, and direct child assessments

Age Range and Administration Interval:

For child-related questions and assessments, prenatal through 36 months.

For parent-related questions, all adults.

Personnel, Training, Administration, and Scoring Requirements:

The requirements vary, depending on the particular measure. Some may be completed by parents and scored by a person with minimal training. Others must be administered and scored by a highly trained individual.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 and higher)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (some measures require administration and scoring by a highly trained individual)

Description: The Parent Interviews (PIs), Father Interviews (FI), and Child Assessments (CA) developed for the national Early Head Start Research and Evaluation Project were designed to include instruments that assess potential program effects on a variety of domains. The instruments included in each data collection interview/assess-

ment were drawn from a variety of sources and include published instruments (many of them are described in individual entries in this resource guide), questions drawn from other large national surveys, and questions that were developed specifically for this study.

The assessments used in the evaluation are summarized in table format at www.mathematica-mpr.com/3rdLevel/ehstoc.htm. The table in Appendix C lists key child and family measures selected for the evaluation. In addition to scales and standardized tests, the interviews and assessments included a number of single items that are simple to administer and use for comparison with the national evaluation results. These include questions about bedtime routines, reading to children at bedtime, frequency of reading to children, and spanking. The interviews can be found at the web addresses listed above.

The results of the evaluation (through age 3) are included in two reports and their appendices and are available at www.mathematica-mpr.com and www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_instruments.html.

Uses of Information: The Early Head Start interviews and assessments can be used by programs to obtain a wide range of parent, child, and service use information useful for performance measurement that can be compared to the national study findings. For copyrighted instruments, programs must obtain permission to use the assessments and must pay for their use.

Reliability: The technical appendices of the two reports include internal consistency reliability for all of the summary scores. As a general rule, summary scores were not included in the report if their reliability was not above .65. The single item questions do not require computing summary scores.

Validity: The assessments were included in the evaluation because they had been used before in large studies and had demonstrated construct validity. Validity work based on the data collected was not reported in the two reports.

Method of Scoring: Each assessment is scored according to the rules and advice from the assessment developers or publishers. Some required complex computer scoring programs and others could be scored by hand and were a simple percentage. Scoring procedures for each measure are summarized in the reports. The single item questions do not require computing summary scores.

Interpretability: Some of the assessments were easily interpretable, while others required a well-trained individual.

Training Support: As part of the evaluation project, in-depth training manuals were developed; these can be obtained by requesting them from Jackie Allen at jallen@mathematica-mpr.com. In addition to the manuals, interviewers and assessors attended a central training session and had to meet rigorous standards before administering the study instruments. Mathematica is not providing any training support for the measures.

Adaptations/Special Instructions for Individuals with Disabilities: Contact Mathematica for more information about how the protocols were adapted for use with individuals with disabilities.

Report Preparation Support: None described.

References:

ACYF. "Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report." Washington, DC: Administration on Children, Youth and Families, 2001. www.mathematica-mpr.com or http://www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_reports.html.

ACF. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC:

Administration for Children and Families, June 2002. www.mathematica-mpr.com or http://www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_reports.html.

Sprachman, S., C. DeSaw, L. Mendenko, M. Salem, K. Boller, and B. Kolln. "Early Head Start National Evaluation Data Collection Training Manual for 6-Month Parents Services Interview, 14-Month Parent-Child Interview and Assessment, and Child Care Observations." Princeton, NJ: Mathematica Policy Research, Inc., December 1996.

www.mathematica-mpr.com

EARLY LEARNING ACCOMPLISHMENT PROFILE REVISED EDITION (E-LAP), 1995

Authors:

M. Elayne Glover, Jodi L. Preminger, Anne R. Sanford

Publisher:

Kaplan Press
(800) 334-2014

Initial Material Cost:

Early Learning Accomplishment Profile (E-LAP) Kit:
\$335: Includes E-LAP Scoring Booklets, E-LAP
Manual, and E-LAP Activity Cards, and Content and
Overview Video.
Demonstration Video (training) costs \$66.

Representativeness of Norming Sample:

Research sample was representative of Year 2000 based on the 1995 US Census Bureau population projection. Sample included 285 children ages 2 to 44 months old, including children with typical and atypical development. A stratified sampling procedure was used based on geographic region, age, race, gender, and type of setting. Children were selected from child care centers, Early Head Start programs, and individual homes. The sample included a representative percentage of children with disabilities (reflecting the U.S. rates for children with disabilities – U.S. Census Bureau, 1995). The assessments were conducted between November 1999 and July 2000.

Languages:

English and Spanish

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 36 months

Personnel, Training, Administration, and Scoring Requirements:

In addition to having knowledge about the target population, a 2-day training workshop is suggested prior to using the E-LAP. It takes an experienced examiner at least an hour to administer the E-LAP and approximately 10 minutes per domain to score

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher for internal consistency, test-retest, and interrater reliability)

Validity: 3 (.5 or higher for concurrent validity)

Norming Sample Characteristics: 3 (nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The E-LAP is designed to help assess overall development of children with special needs functioning, based on developmental

milestones focusing on the birth through 3-year age range. The E-LAP is a criterion-referenced tool that focuses on the following domains: Gross

Motor, Fine Motor, Cognitive, Language, Self-Help, and Social/Emotional skills. Items in the E-LAP were drawn from various early childhood assessment instruments.

Uses of Information: The E-LAP provides guidance to early childhood programs in assessment and programming for infants, young children, and children with special needs.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) ranges from .84 to .98 for Gross Motor, with a total of .99; .90 to .96 for Fine Motor, with a total of .98; .96 to .97 for Cognitive, with a total of .99; .91 to .96 for Language, with a total of .98; .93 to .97 for Self-Help, with a total of .98; and .87 to .91 for Social Emotional, with a total of .96. (2) Test-retest reliability was measured one to three weeks apart only for a small subsample of children (92). Correlations for the domains ranged from .96 to .99. (3) Inter-rater reliability correlation coefficients for the domains ranged from .96 to .99.

Validity: (1) Concurrent validity was tested using the Mental and Motor Scales of the BSID-II (Bayley Scales of Infant Development). Results indicate a strong correlation (.90 to .97) between the E-LAP and BSID-II scored in each domain for the overall sample. In the 2- to 12-month sample, correlations ranged from .83 to .95, and in the 13- to 24-month sample, correlations ranged from .72 to .88. Correlations in the 25- to 36-month sample were lower, ranging from .47 to .83. The manual mentions that this lower correlation, in addition to other analytical information, may indicate that the ELAP is a less effective assessment tool

for older children.

Method of Scoring: The examiner must first calculate the child's chronological age (by months) to determine the appropriate starting point in each domain. Each item is marked with a plus (+) if the child exhibits the criterion-referenced behavior or a minus (-) if the skill is not demonstrated by the child. Examiners must establish a basal (8 consecutive items successfully completed) and a ceiling (3 errors out of 5 consecutive items). All items prior to the basal are counted as correct. The raw score represents these items plus the number of items successfully completed in the domain up to the ceiling. The manual provides further guidelines for computing the raw score for the domain, and calculating the corresponding developmental age-range that the child falls within. It also suggests that any modifications of the procedures or use of adaptive equipment be included in the comments section of the scoring booklet in order to better understand a child's skills. The Content and Overview video also includes information on scoring the E-LAP.

Interpretability: E-LAP is a non-standardized test, and should be used in conjunction with norm-referenced assessments to determine whether or not a child has a disability. In computing scores, the manual also states that it is important for the examiner to be aware that the normative developmental age assigned to a specific item varies among research sources, and that developmental ages need to be viewed as approximate. Although the manual does not provide

normed scores, it does provide means and standard deviations for the scores obtained in the various domains for the project sample (including both typical and atypical children) as well as the core sample (including only children with presumed typical development).

Training Support: There is a demonstration video available to assist transdisciplinary teams with using the E-LAP. There are also individuals from the Chapel Hill Training-Outreach Project available for training on the E-LAP. Information on training can be obtained by calling 800-334-2014, ext.5100.

Adaptations/Special Instructions for Individuals with Disabilities: The manual suggests that in the case of children with disabilities, reports of diagnostic results should be used to provide information regarding the child's developmental level of functioning (to help determine the appropriate point for beginning the assessment process). If that information is not available, the manual suggests that the assessor begin administering the E-LAP at half of the child's

chronological age, which would allow for the establishment of the basal. The manual assumes that the teacher will create appropriate developmental milestones for children with more involved disabilities, and if necessary, should appropriately modify these milestones into sub-objectives for the child.

Report Preparation Support: The information in the E-LAP recording procedures suggest that the absence of certain skills (as demonstrated through the assessment) should be incorporated into the child's Individualized Family Service Plan (IFSP).

References:

Glover, Elayne M., Jodi L. Preminger, and Anne R. Sanford. Early Learning Accomplishment Profile Revised Edition (E-LAP). Lewisville, NC: Kaplan Press, 1995.

Hardin, Belinda J., and Ellen S. Peisner-Feinberg. The Early Learning Accomplishment Profile (Early LAP) Examiner's Manual and Reliability and Validity Technical Report. Lewisville, NC: Kaplan Press, 2001.

EYBERG CHILD BEHAVIOR INVENTORY (ECBI) AND SUTTER-EYBERG STUDENT BEHAVIOR INVENTORY-REVISED (SESBI-R), 1999

Authors:

Sheila Eyberg and Donna Pincus

Publisher:

Psychological Assessment Resources
(800) 331-TEST
www.parinc.com

Initial Material Cost:

Contact publisher for cost information

Representativeness of Norming Sample:

The ECBI is standardized on socio-economically diverse Caucasian families and families of other ethnicities with an equal representation of children between the ages of 2 and 16 years. The SESBI-R was originally standardized on lower-middle socio-economic status (SES) preschoolers. Caution is given regarding the interpretation of cutoff scores in different geographical regions.

Languages:

English ¹

Type of Assessment:

Direct child assessment and parent report

Age Range and Administration Interval:

2 to 16 years

Personnel, Training, Administration, and Scoring Requirements:

Tests can be administered and scored by individuals who do not have clinical training. Completion of the forms requires at least a 6th grade reading level. Training requires familiarizing oneself with the manual and questionnaires. The ECBI and SESBI-R require 10 minutes each to complete and under 5 minutes each to score. Interpreting the scores requires graduate training in psychology, counseling, or a closely related field.

Summary

Initial Material Cost: not available

Reliability: 3 (.65 or higher)

Validity: 2 (under .5 for concurrent validity and under .4 for predictive validity)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (self-administered and scored by someone with basic clerical skills, but interpreted by a highly trained individual)

¹Unofficially translated (by universities and other organizations) into Welsh, Australian, Spanish, German, Norwegian, French, and Chinese.

Description: The ECBI and SESBI-R are rating scales that assess the severity of conduct problems in children ages 2 through 16 years as well as the extent to which parents and teachers find the behaviors troublesome. The ECBI, which consists of 36 items, is completed by parents and assesses the frequency of disruptive behaviors occurring in the home. The SESBI-R, which consists of 38 items, is completed by teachers and is useful in the assessment of disruptive behaviors in the school setting. Each test provides an Intensity Raw Score and a Problem Raw Score.

Uses of Information: Taken together, the ECBI and SESBI-R have multiple applications, including use as (1) screening measures in the clinical identification of children for the diagnosis and treatment of externalizing behavior problems, (2) screening measures in the identification of “high-risk” children for delinquency prevention programs, and (3) outcome measures in evaluations of treatment programs for conduct-disordered children.

Reliability: (1) Internal consistency reliability (Cronbach’s alpha): for the ECBI, .95 for the Intensity scale and .93 for the Problem scale. For the SESBI-R, .98 for the Intensity scale and .96 for the Problem Scale. The Intensity scale had a mean item-to-total correlation of .76, and the Problem scale had a mean item-to-total correlation of .65. (2) Test-retest reliability: for the ECBI, .75 to .86 for the Intensity scale and .75 to .88 for the Problem scale. For the SESBI-R, .87 for the Intensity scale and .93 for the Problem scale. (3) Inter-rater reliability: for the ECBI, .86

for the Intensity scale and .79 for the Problem scale. For the SESBI-R, inter-rater reliability for the Intensity scale was in the .85 to .86 range for the Intensity scale and was .84 to .87 for the Problem scale.

Validity: (1) Concurrent validity: for the ECBI, Problem and Intensity scores were significantly correlated with both the Externalizing scale (.67 and .75, respectively) and the Internalizing scale (.48 and .41, respectively) of the Child Behavior Checklist. Also, the ECBI was found to correlate significantly with the Parenting Stress Index (PSI): the ECBI Problem and Intensity scores were significantly correlated with the PSI Child Domain scores (.45 and .45, respectively). The scores obtained on the SESBI, SESBI-R, and the Revised Edition of the School Observation Scale (REDSOCS) in the regular classroom were significantly related to off-task and inappropriate behavior categories, but not to noncompliance. (2) Predictive validity: For the SESBI-R, correlations between scores and both the child’s number of school suspensions and the number of referrals to the school principal for conduct problems were .26 to .39 one year later and .21 to .36 two years later.

Method of Scoring: Each behavior is rated on two scales: a 7-point Intensity scale assesses how often the behaviors currently occur in the home or school setting (1 means “never,” 4 means “sometimes,” and 7 means “always”) and a Problem scale (Yes/No) identifies whether the child’s behavior is problematic for the parent or teacher. Scores are computed by summing the

Intensity scale scores on each page for pages 1 and 2. Similarly, the number of “Yes” responses are summed to come up with a separate Problem scale total for pages 1 and 2. To obtain the Intensity Raw score, the Intensity scale scores from pages 1 and 2 are summed. To obtain the Problem raw score, the Problem scale scores from pages 1 and 2 are summed.

Interpretability: Both the ECBI and SESBI-R are continuous in that higher scores on the scale indicate a greater level of conduct-disordered behavior and a greater impact on the parent or teacher. Comparison with normative data in Chapter 3 of the manual allows for more specific examination of the significance of scores, especially in cases where the cutoff score of either scale has been exceeded. T-score conversions for the raw scores are provided in the appendices of the manual (Appendices C, D, E, and F). T-score conversions for either of the scales that are greater than or equal to 60 are clinically significant. T-Scores below 60 are within the normal range. Caution is issued when interpreting SESBI-R cutoff scores because the author’s analyses with the SESBI-R suggest that for the kinds of behaviors assessed, there is wide variability both within and between geographic locations. Therefore, clinicians should adjust cutoff levels based on local norms as needed.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Two case studies are presented in the examiner’s manual.

References:

Eyberg, Sheila, and Donna Pincus. The ECBI & SESBI-R: Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual. Odessa: Psychological Assessment Resources, 1999.

FUNCTIONAL EMOTIONAL ASSESSMENT SCALE (FEAS), 2001

Authors:

Stanley Greenspan, Georgia DeGangi and Serena Wieder

Publisher:

The Interdisciplinary Council on Developmental and Learning Disorders
www.icdl.com

Initial Material Cost:

Text Book: \$40 for ICDL members, \$47 for non-members
Additional protocol booklets: \$8

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Direct observation and possible direct child assessment

Age Range and Administration Interval:

7 months to 4 years (research version); Six versions: 7-9 months; 10-12 months; 13-18 months; 19-24 months; 25-35 months; 3-4 years

Personnel, Training, Administration, and Scoring Requirements:

Takes 15-20 minutes to administer. The examiner should be trained and experienced. The authors recommend videotaping the caregiver-child play interaction session. They advise that live scoring should not be attempted without first observing at least 10 videotapes with at least an 80 percent reliability in scoring live and videotaped observations.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher for inter-rater reliability; no other reliability provided)

Validity: 2 (<.5 for concurrent)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Functional Emotional Assessment Scale (FEAS) provides a framework for observing and assessing a child's emotional and social functioning in the context of the relationship with his or her caregiver as well as the caregiver's capacity to support the child's emotional development. The FEAS assesses the child

on six levels of social and emotional development: (1) regulation and interest in the world, (2) forming relationships (attachment), (3) intentional two-way communications, (4) development of a complex sense of self, (5) representational capacity and elaboration of symbolic thinking, and (6) emotional thinking or development and

expression of thematic play. There are two versions of the FEAS, a clinical version and a research version. The research FEAS, which evolved from the clinical FEAS, has cutoff scores to assist in interpreting the results and has been used to test for the scale's validity and reliability. Each of these has versions that are designed for different age groups. In both versions, the caregiver (parent) is asked to play with his or her child as he/she might at home for 15 minutes with 3 different types of developmentally appropriate toys: symbolic toys, tactile toys, and toys involving large movement activities. The examiner may also want to engage the child in play to attempt to elicit behaviors not observed during the caregiver-child play interaction. Because considerable experience is needed to score the FEAS reliably in live observation sessions, the authors recommend that these unstructured play observations be videotaped and scored later. The scale should be used in conjunction with other instruments as part of an overall assessment.

Uses of Information: The FEAS is intended to help clinicians identify critical areas deserving of further clinical inquiry. It can be used descriptively to profile children's emotional, social, and related developmental capacities. It can also be used to diagnosis or screen for problems in children who are experiencing regulatory disorders, but not to formally diagnose specific disorders.

Reliability: (1) Inter-rater reliability (Cronbach's alpha): The alpha coefficients between pairs of observers viewing between 15 and 46 videotaped caregiver-child interactions ranged from .90 to .92 for the caregiver scale and .90 to .98 for the child scale. The alphas between a pair of observers viewing 15 interactions, one coding the interactions live and the other a videotape of the interactions, were .83 for the caregiver scale and .89 for the child scale.

Validity: Four non-nationally representative samples of young children between the ages of 7 and 48 months, except when noted otherwise, were used to test for validity: (1) 197 normal children; (2) 190 children with regulatory disorder; (3) 41 children between the ages of 19 and 48 months with pervasive developmental disorder; and (4) 40 multi-problem children. All of the samples had a larger proportion of boys, white, and middle-class children. (1) Construct validity:¹ The scores obtained by normative and clinical samples of young children were compared using a discrimination index, t-tests, and analysis of variance. (2) Accuracy of cutoff scores (ranges for the different age groups): False normal errors for the total (child and caregiver) scale ranged from 5 to 28 percent, false delay errors ranged from 26 to 63 percent, specificity (probability correctly identifying a normal child) ranged from 37 to 74 percent, and sensitivity (probability of correctly identify-

¹Note that although the authors consider this information to reflect construct validity, the relationships described are consistent with the way concurrent validity is used throughout this resource guide.

ing a delayed child) ranged from 74 to 95 percent. (3) **Concurrent:** Intercorrelations between the FEAS scores during symbolic and tactile play and two other instruments developed by the authors, the Test of Sensory Functions in Infants and the Test of Attention in Infants, were not significant. The authors interpret this to mean that the FEAS provides unique information.

Method of Scoring: The clinical FEAS may be left unscored and used to provide a descriptive profile of the young child's developmental capacities or to help systematize clinical thinking. The scale can also be used to rank each item as follows: capacity not present (0), capacity fleetingly present (1), capacity intermittently present (2), capacity present most of the time (3), capacity present all of the time in all circumstances (4), or no opportunity to observe capacity (not applicable). The ratings can then be summed for each functioning area and divided by the functioning area's maximum possible score to obtain a percentage. However, only some of the developmental functioning areas can be described quantitatively; the others should be described qualitatively. The research FEAS rates both the caregiver and the child on their mastery of the skill as follows: behavior is not seen or is observed only briefly (skill not mastered) (0), behavior is present some of the time or observed several times (skill partially mastered) (1), and behavior is consistently observed or observed many times (skill mastered)

(2). The ratings can be summed to obtain category and subtest scores for the caregiver and category and subtest scores for the child, as well as a combined caregiver total score and a combined child total score.

Interpretability: A developmental growth chart can be used to help assess the child's functional developmental accomplishments over time based on information collected from the clinical FEAS. The research FEAS has cut-off scores that can be used to determine if parent-child interaction patterns are normal, at risk, or deficient. However, in interpreting both the clinical and research FEAS, the authors strongly recommend that the FEAS not be used alone, but as part of a comprehensive assessment of the caregiver-child relationship.

Training Support: The author offers a training course for the FEAS. Information can be acquired by calling 301-320-6360 or by visiting www.icdl.com

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

DeGangi, G. *Pediatric Disorders of Regulation in Affect and Behavior. A Therapist's Guide to Assessment and Treatment.* San Diego, CA: Academic Press, 2000.

HAWAII EARLY LEARNING PROFILE (HELP), 1997

Authors:

Stephanie Parks

Publisher:

VORT Corporation
(650) 322-8282
www.vort.com

Initial Material Cost:

Inside HELP: Administration and Reference Manual, \$50
HELP Family-Centered Interview (0-3), \$25 (package of 25)
HELP Strands (0-3), \$3 each
HELP Checklist (0-3), \$3 each
HELP Charts (0-3), \$3 each
HELP Activity Guide (0-3), \$28 each

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Direct child or parent assessment

Age Range and Administration Interval:

0 to 36 months

Personnel, Training, Administration, and Scoring Requirements:

HELP is for use by trained professionals. An initial direct assessment may last from 45 to 90 minutes, but in some cases may be completed in 15 to 20 minutes. Most of the scoring is done during the assessment.

Summary

Initial Material Cost: 1 (<\$100)
Reliability: 1 (none described)
Validity: 1 (none described)
Norming Sample Characteristics: 1 (none described)
Ease of Administration and Scoring: 3 (administered and scoring by a trained professional)

Description: HELP is a curriculum-based assessment consisting of 685 developmental skills and behaviors covering six traditional child developmental domains: cognitive, language, gross motor, fine motor, social-emotional, and self-help. The developmental skills are organized by skill domains and, within skill domains, by “conceptual strands.” Within each strand, the skills are sequentially ordered by age. For the assess-

ment, the examiner selects several developmental skills above and below the child’s approximate developmental age from each developmental domain and prepares 5 to 10 play or daily activities to elicit several skills concurrently. In addition to working with the child using the prepared activities, the examiner also observes the child’s environment and interviews the parents for information.

Uses of Information: HELP can be used to identify needs, track growth and development, and develop plans for meeting objectives. It can be used to identify a child's developmental skills and behaviors along multiple lines of development, the child's developmental skill and behavior strengths and needs, physical and social environment factors affecting development, and the way development in one area may be affecting development in other areas.

Reliability: None reported.

Validity: Face and content validity tests were used to select and group the skills. The skills were selected from growth-and-development scales and standardized tests. An interdisciplinary team of pediatric therapists grouped the skills into strands and sequentially ordered them by age. No tests of concurrent or predictive validity reported.

Method of Scoring: Behaviors are scored as "present," "not present," "emerging," "atypical/dysfunctional," or "not applicable." HELP Strands or HELP Checklist can be used to record outcomes and track progress. Because it divides each domain into strands, HELP Strands is recommended for children with disabilities or more uneven development within a domain.

Interpretability: Although the manual does not provide any exact rules or formulas for deter-

mining approximate developmental levels, it does provide general "rule of thumb" guidelines. The manual provides an explanation for a delay or atypical behavior for each skill area, along with appropriate interventions.

Training Support: A 20-minute training video, *Using HELP Effectively*, can be purchased from VORT for \$14.95. It is recommended that this video be used in conjunction with the *Inside HELP*. The instruction chapter of the *Inside HELP* manual can be found in the training section on the VORT web page.

Adaptations/Special Instructions for Individuals with Disabilities: *Inside Help* has examples of adjustments to accommodate specific disabilities and special needs when assessing skills and behaviors.

Report Preparation Support: HELP Strands and HELP Checklist can be used to communicate the child's progress to parents.

References:
Parks, Stephanie. *Inside HELP: Administration and Reference Manual*, 1997 Revision. Palo Alto, CA: VORT Corporation, 1999.

VORT Corporation. *Using HELP Effectively*. Palo Alto, CA: VORT Corporation, 1994

HIGH/SCOPE CHILD OBSERVATION RECORD (COR) FOR AGES 2 1/2-6, 1999

Authors:

High/Scope Education Research Foundation

Publisher:

High/Scope Press
(313) 485-2000
press@highscope.org

Initial Material Cost:

COR Kit: \$125 (includes a Manual, 25 Assessment Booklets, 4 sets of Anecdotal Notecards, 50 Parent Report)

Representativeness of Norming Sample:

No norming sample described.

Languages:

English and Spanish

Type of Assessment:

Observation

Age Range and Administration Interval:

2 1/2 to 6 years; assessment is intended to be a full-year assessment usually done two to three times throughout the school year.

Personnel, Training, Administration, and Scoring Requirements:

Training in COR for teachers and teaching assistants is recommended. For administering, the manual recommends focusing on a few children each day or two and writing notes on those children specifically. If done by hand, scoring takes approximately one hour. If done on computer, scoring is less than five minutes.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (majority of correlations are <.50)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic skills)

Description: With the COR, a trained teacher or observer assesses each child's behavior and activities in six categories of development: (1) initiative, (2) social relations, (3) creative representation, (4) music and movement, (5) language and literacy, and (6) logic and mathematics. Over several months, the teacher writes brief, anecdotal notes describing examples of children's

behavior in these six categories. The teacher then uses these notes to rate the child's behavior on 30 five-level COR items within these categories. The COR can be administered at various points throughout the year to measure change over time, or at a single point in time to measure the current developmental level of a child.

Uses of Information: The COR can be used

to assess the educational progress of individual children or a group of children as a whole and the program's curriculum's contribution to children's development. It can also be used to develop program plans that focus on specific areas of child development based on the outcomes of the assessment at the individual level. The results can also be shared with the next year's program staff as well as parents.

Reliability: (1) Internal consistency (Cronbach's alphas): ranged from .80 to .93 for teachers and .72 to .91 for assistant teachers. (2) Inter-observer reliability (Pearson correlation coefficients): ranged from .61 to .72.

Validity: (1) Concurrent validity: correlations of the COR development categories with similar categories on the McCarthy Scales of Children's Abilities ranged from .27 to .53. The correlations between the COR and all of the McCarthy Scale categories ranged from .27 to .66.

Method of Scoring: Items are scored by the six categories. Using the anecdotal notes, the highest level of behavior that is characteristic of the child is checked. The levels go from one through five, with five being the highest score possible for each item. If scoring by hand, there is a formula to use; if scoring on the computer, the computer does it for you.

Interpretability: The COR is meant to follow a child's (or children's) development over time, for instance, over the school year.

Training Support: The COR manuals have written support with examples. However, High/Scope recommends that teachers participate in a two- or three-day workshop on the use of COR offered throughout the country. The training covers how to recognize developmentally significant behavior and describe it in anecdotal notes, how to select the item and item level that each anecdotal note represents, and how to report these results to parents and program officials. The training is \$190 per person for two days, and \$95 per person for the additional computer-training day. Also, those individuals who go through training are given a practice CD, which is used as a follow-up/refresher once training has been completed. Training information and a schedule is available on the High/Scope web page, www.high-scope.org/TrainingConferences/homepage.htm.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Parent Report Forms are included in the package and are used as the basis for discussion at parent conferences.

References:

High/Scope Educational Research Foundation. High/Scope Child Observation Record (COR) For Ages 2 1/2-6. Ypsilanti, MI: High/Scope Press, 1992.

HIGH/SCOPE CHILD OBSERVATION RECORD FOR INFANTS AND TODDLERS (COR-IT), 2002

Authors:

High/Scope Educational Research Foundation

Publisher:

High/Scope Press

(800) 407-7377

www.highscope.org

Initial Material Cost:

COR-IT Kit (includes 3 Observation Items Booklets, 1 User Guide, 25 Anecdote Forms, 25 Observations About Your Child Forms, 25 Parent Guides, 25 Child Information and Development Summary Forms, 5 Group Summary Forms, and posters): \$150

Representativeness of Norming Sample: None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

6 weeks to 3 years. Administer 2 to 3 times during a program year.

Personnel, Training, Administration, and Scoring Requirements:

No special degrees are required to conduct the observation, however, it is recommended that the observer receive a two-day training by High/Scope and be in regular contact with the children being observed. The observer should do the observation over a period of several weeks or months and spend 30 to 60 minutes writing rough notes and transforming them into formal notes.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 and higher for both internal and inter-rater reliabilities)

Validity: 3 (.5 and higher for concurrent validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The High/Scope Child Observation Record for Infants and Toddlers (COR-IT) enables the user to conduct an ongoing, comprehensive, systematic assessment in programs serving children from the ages of 6 weeks to 3 years. COR-IT's 28 items are divided into 6 categories: (1) sense of self, (2) social rela-

tions, (3) creative representation, (4) movement, (5) communication and language, and (6) exploration and early logic. Over weeks or months, the caregiver records brief, anecdotal notes describing examples of children's behavior in these six categories. At the end of the observation period, which usually takes place over a period of weeks

or months, the caregiver uses these notes to complete a development summary form. The COR-IT has a computer software version that allows individuals to use their computers to record and store observations and to generate COR scores and reports based on this information.

Uses of Information: COR-IT provides an accurate assessment of a child's development and abilities, which can be used to help programs plan activities for the child, to monitor the effects of these activities and the progress of the child, to assess the effectiveness of the program and its curriculum, and to provide a framework for communicating with parents about a child's developmental needs.

Reliability: (1) Internal consistency reliability (Cronbach alpha): The alpha for the entire 28-item scale was .99, and the alphas for the six categories were .92 or .93. The alphas for the 28-item scale for each age category were .94 for infants under 1 year, .95 for children 1 to 2 years, and .78 for children 2 to 3 years. (2) Inter-rater reliability: The Pearson correlations between two groups of observers were .93 for the overall scale and ranged from .83 to .91 for the six categories.

Validity: (1) Concurrent: The correlations between COR-IT and the Bayley Mental and Motor Age Scores were .87 and .91, respectively. The correlations ranged from .88 to .92 between the six COR-IT categories and the motor age score and .83 to .90 between the categories and the mental age score. Since both the COR-IT and the Bayley scores are strongly influenced by the child's age, the authors also tested for validity with

the effects of age statistically removed from the correlations. When they did this, they obtained correlations between the COR-IT scale and the mental and motor scores of .26 and .36 respectively.

Method of Scoring: The Observation Item manual provides 5 examples of typical behavior for each of the 28 items. For each item, the caregiver compares those examples with his/her notes to rank the child's typical behavior of the item on a five-point scale from simple (1) to more complex (5) in the development summary form and also enters the highest level of behavior the child achieved. The form provides instructions to compute the average and composite scores. If the program requires a group summary, the caregiver completes the group summary form using the information on the children's development summary form and follows the instructions on the form to compute the average and group growth scores.

Interpretability: The results from the observations are intended to be used by both the caregiver and the program administrators. No descriptions are provided on how to interpret the results.

Training Support: High/Scope recommends the caregiver attend a two-day workshop on the use of COR-IT, which is offered throughout the country. There is also a one-day training on the computer software. Information on the training schedule and topics are available on the High/Scope web site or by emailing: training@highscope.org.

Adaptations/Special Instructions for Individuals with Disabilities: Adaptations and special instructions are unnecessary as long as the child's level of functioning is between ages 6 and 36 months. However, High/Scope cautions that the reliability and the validity of the instrument with special needs populations have not been established.

Report Preparation Support: An Observation About Your Child form can be used to prepare a report for parents. The COR-IT computer software allows for presentation of ratings in both graphic and narrative form.

References:

High/Scope Educational Research Foundation. High/Scope Child Observation Record for Infants and Toddlers User Guide. Ypsilanti, MI: High/Scope Press, 2002.

HUMANICS NATIONAL INFANT-TODDLER ASSESSMENT, 1994

Authors:

Jane A. Caballero and Derek Whordley

Publisher:

Humanics Psychological Test Corp.

(800) 874-8844

(404) 874-1976 (fax)

www.humanicslearning.com

info@humanicslearning.com

Initial Material Cost:

Manual: \$19

Forms (25): \$35

Representativeness of Norming Sample:

None. The authors recommend that each facility develop local norms and provides the formula to derive percentage rank.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 3 years; four-month intervals are recommended between assessments, but the examiner can choose to deviate from this.

Personnel, Training, Administration, and Scoring Requirements:

Can be administered by teachers or parents. Training to enhance observational skills is desirable, and some knowledge of child development concepts is required to understand the importance of assessment items.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic clerical skills)

Description: The Humanics National Child Assessment Form (HNCAF) is a checklist of 90 skills and behaviors exhibited by children during the first three years of life. The skills/behaviors are grouped into four broad categories—social-emotional, language, cognitive, and gross and fine motor development. The assessment lists 18 skill/behavior items in each of these five areas, listed in the order they are likely to emerge. The assessment can be administered either through informal observations or through activities with

the child.

Uses of Information: HNCAF is designed for use by caregivers to identify the skills and behavior a child has, to screen for developmental or physical impairments, to plan learning experiences that facilitate further growth, and to monitor the child's progress. HNCAF can also be used for parent training.

Reliability: Not discussed.

Validity: Not discussed.

Method of Scoring: The child's responses are scored by checking one of two boxes (occurs consistently or occurs occasionally) or, if the behavior is not present, by leaving the boxes unchecked.

Interpretability: The manual discusses how the assessment results can be used to develop an individual profile and educational plan for the child and has an example of how this might be done.

Training Support: Not mentioned in the manual. The authors suggest reading Betty Rowen's book, *The Children We See* (Holt, Rinehart and Winston, Inc., 1973) for techniques for observing children's behavior.

Adaptations/Special Instructions for Individuals with Disabilities: The manual describes areas of disabilities and indicators of these disabilities that may be observed while administering the assessment.

Report Preparation Support: The manual provides an example of a completed report.

References:

Caballero, Jane and Derek Whordley.
Humanics National Infant-Toddler Assessment Handbook: A User's Guide for the Humanics National Child Assessment For Ages 0-3. Atlanta, GA: Humanics Psychological Test, Corp., 1994.

INFANT-TODDLER DEVELOPMENTAL ASSESSMENT (IDA), 1995

Authors:

Sally Provence, Joanna Erikson, Susan Vater, and Saro Palmeri

Publisher:

Riverside Publishing
(800) 323-9540
www.riverpub.com

Initial Material Cost:

IDA Complete Kit: \$502 (includes 25 Parent Reporting Forms, Health Record Guides, and Record Forms, a Foundations and Study Guide, the Administration Manual, readings, and IDA Manipulative Kit in Carrying Case)

Without Manipulative Kit and Carrying Case: \$278

Representativeness of Norming Sample:

The research sample of 100 children between birth and 3 years old is not nationally representative.

Languages:

English¹

Type of Assessment:

Parent report and observation

Age Range and Administration Interval:

Birth to 42 months

Personnel, Training, Administration, and Scoring Requirements:

A multidisciplinary team or a very well-trained clinician can administer the assessment. Training tapes are available. Administration and scoring times vary.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher)

Validity: 3 (based on percent scoring agreement with other instruments)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Infant-Toddler Developmental Assessment (IDA) is designed to improve early identification of children birth to 3 years of age who are developmentally at risk. There are six IDA phases that are designed to be conducted by a team of two or more professionals: (1) Referral & Pre-interview Data Gathering, (2) Initial Parent Interview, (3) Health Review,

(4) Developmental Observation and Assessment, (5) Integration and Synthesis, and (6) Share Findings, Completion, and Report. Each phase develops from the preceding one and is completed only after team discussion and review. The Developmental Observation and Assessment Phase (Phase Four) uses the Provence Birth-to-Three Developmental Profile, which provides a

¹ The parent report is also available in Spanish.

descriptive summary of a child's developmental competencies. The entire assessment uses observation and parent reports of the child's development along eight developmental domains: Gross Motor, Fine Motor, Relationship to Inanimate Objects (Cognitive), Language/Communication, Self-Help, Relationship to Persons, Emotions and Feeling States (affects), and Coping. The Provence Profile is to be used within the context of the full IDA rather than as an isolated test. Five forms are used to gather and record information: Parent Report, Request for Health Information, Family Recording Guide, Health Recording Guide, and IDA Record.

Uses of Information: The IDA helps determine the need for monitoring, consultation, intervention, or other services for the child and family and may be used to develop an Individualized Family Service Plan (IFSP).

Reliability: (1) Internal consistency reliability: alpha coefficients for the Provence domain scores range from .90 to .96 for ages 1 to 18 months and .77 to .96 for ages 19 to 36 months. (2) Test-retest reliability: no information available. (3) Inter-rater reliability: correlations between raters ranged from .91 to .95 for seven of the eight domains and .81 for the remaining domain (language/communication).

Validity: (1) Concurrent validity: comparisons between the IDA and the Bayley Scales of Infant Development, Hawaii Early Learning Profile, Learning Accomplish Profile, and the Vineland Adaptive Behavior Scales showed that, of the items IDA had in common with those instru-

ments, the percentage agreement on the developmental age ranged from 84 to 100 percent. A comparison of the IDA with the Bayley and Vineland Adaptive Behavior Scales in identifying children needing services found the IDA and Vineland had more similar outcomes. The IDA classified 51.6 percent of the children as needing services compared to 66.0 to 73.6 percent for the Vineland and 13.2 to 22.6 percent for the Bayley. Of 57 children referred to services by IDA practitioners, according to the receiving agencies, 83 percent of the referrals were considered to be appropriate. (2) Predictive validity: no information available.

Method of Scoring: The Provence Protocols behavioral items are marked "present and observed," "not present or observed," "reported present and not observed," "reported not present," "emerging," or "refused." The number of correct responses is used to determine the child's performance age, which is compared to the child's chronological age to determine whether to rate the child's development in the domain as "competent" or "of concern" and, if the latter, the degree of the delay. Using tables in the manual, the "Percentage Delay" can also be computed from the child's observed performance age and the child's chronological age, adjusted for prematurity.

After obtaining the Provence Protocol score, the assessment team needs to take into account qualitative aspects of the child's performance and performance on certain "marker" skills for the child's age group to decide again whether the child's development is "competent" or "of con-

cern” and, if the latter, the level of concern.

Interpretability: The manual provides a general guideline on how to interpret the results and what should follow based on the results.

Training Support: Training materials are available to help professionals train others on the administration of the IDA. Materials consist of a Leader's Guide and three videos. These materials are not meant to be a "self-study" course for individuals trying to learn IDA. The Erikson Institute also provides training. For more training options, visit www.erikson.edu/.

Adaptations/Special Instructions for Individuals with Disabilities: IDA meets the cri-

teria for assessment as required by the regulations for the Individuals with Disabilities Education Act (IDEA).

Report Preparation Support: The Manual gives instructions on how to share findings and develop a plan with parents in Phase Six, “Share Findings, Completion, and Report.”

References:

Provence, S., J. Erikson, S. Vater, and S. Palmeri. Infant-Toddler Developmental Assessment (IDA) Administration Manual. Itaska, IL: Riverside Publishing, 1995.

www.erikson.edu/

INFANT TODDLER SOCIAL AND EMOTIONAL ASSESSMENT (ITSEA), 2001

Authors:

Alice S. Carter and Margaret J. Briggs-Gowan

Publisher:

Contact Alice S. Carter at Alice.Carter@umb.edu, Margaret J. Briggs-Gowan at Margaret.Briggs-Gowan@yale.edu, or the ITSEA project office at ITSEA@yale.edu.

Initial Material Cost:

Free; if the authors give permission, they will email you the forms and manual.

Representativeness of Norming Sample:

Not nationally representative. Age and gender stratified random sample of children born at a Connecticut hospital between July 1995 and September 1997 and who lived in the New Haven-Meridan SMSA. The sample excludes children likely to have significant developmental delays, whose parents lost custody, or who had poor command of English.

Languages:

English, French, Spanish, Hebrew, and Dutch

Type of Assessment:

Parent (or child care provider) report (self-administration) or parent assessment (structured interview)

Age Range and Administration Interval:

1 to 4 years of age

Personnel, Training, Administration, and Scoring Requirements:

The reading level is between 4th and 6th grade. Administration time is 20 to 30 minutes to complete as a questionnaire, and 35 to 45 minutes as an interview. A professional with training in standardized assessment is needed to interpret the results.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (majority is .65 or higher)

Validity: 3 (.5 or higher with CBCL)

Norming Sample Characteristics: 2 (normed within the past 15 years, not representative sample)

Ease of Administration and Scoring: 2 (self-administered, but scored by a trained individual)

Description: The Infant Toddler Social and Emotional Assessment (ITSEA) is designed to detect social-emotional and behavior problems, and delays in the acquisition of competencies in children age 12 to 48 months old. The ITSEA relies on parents and child care providers' observations of the child in natural environments. It can be administered as a questionnaire or a struc-

ured interview. The ITSEA measures four behavioral domains: (1) externalizing using activity/impulsivity, aggression/defiance, and peer aggression scales; (2) internalizing using depression/withdrawal, general anxiety, separation distress, and inhibition to novelty scales; (3) dysregulation using sleep, negative emotionality, eating, and sensory sensitivity scales; and (4) competen-

cies using compliance, attention, imitation/play, mastery motivation, empathy, and prosocial peer relations scales. The ITSEA also includes three indices—maladaptive, atypical behavior, and social relatedness—to identify more serious problems. The indices are comprised of behaviors that have clinical significance but are not necessarily correlated. Items are rated on a three point scale: 0 = not true/rarely, 1 = somewhat true/sometimes, and 2 = very true/ often, with a no opportunity option if the caregiver did not have the opportunity to observe the behavior. There is also a short version, the Brief Infant-Toddler Social and Emotional Assessment, which can be used as an initial screen for the ITSEA.

Uses of Information: The ITSEA is a tool for identifying children and caregivers who may benefit from additional dialogue about children's behavior and development to determine the presence of abnormal behaviors, psychopathology, or delayed competences. By itself, it is not sufficient to make such a determination.

Reliability: (1) Internal consistency (Cronbach's alpha): the domains and scales alphas were .87 for externalizing (.73 to .79 for the scales), .80 for internalizing (.71 to .77 for the scales), .86 for dysregulation (.63 to .84 for the scales), and .90 for competence (.59 to .82 for the scales). The majority of the scales scores exceeded .65. The alphas for the indices were .56 for maladaptive, .56 for social relatedness, and .45 for atypical behaviors. (2) Test-retest reliability (test completed within a 44-day interval): the ITSEA domain and scale coefficients were .82 for exter-

nalizing (.69 to .85 for the scales), .83 for internalizing (.74 to .85 for the scales), .91 for dysregulation (.82 to .88 for the scales), .90 for competence (.77 to .88 for the scales). (3) Inter-rater reliability: agreement between mothers and fathers were .69 for externalizing (.65 to .73 for the scales), .58 for internalizing (.43 to .64 for the scales), .79 for dysregulation (.66 to .78 for the scales), and .76 for competence (.47 to .73 for the scales).

Validity: (1) Criterion and construct validity were evaluated by comparing parent ITSEA ratings with their ratings on the Child Behavior Checklist2/3 (CBCL2/3), Colorado Child Development Inventory (CCTI), Parenting Stress Index (PSI), Mullen Scales of Early Learning, and Vineland Adaptive Behavior Scales. The authors reported that the strength of the relationships between the ITSEA ratings and these other measures varied according to the similarities of the constructs. The correlations between parent reports on their 2-year olds on the ITSEA and CBCL2/3 were .57 and .71 for the internalizing and externalizing domains, respectively. The correlations between the CBCL2/3 Internalizing and Externalizing constructs and ITSEA Dysregulation scores were .52 and .49, respectively. The authors reported that the Maladaptive and Atypical behavior indices correlation with CBCL2/3 Internalizing and Externalizing scores (.15 to .42), as well as the correlations between ITSEA problem domains and PSI domains (between .21 and .45) supported the validity of the ITSEA. According to the authors, other relationships that supported the validity of the ITSEA were the correlations

between ITSEA Internalizing, Externalizing, and Dysregulation problem scales and CCTI Emotionality (between .37 and .53) and Soothability domains (between -.23 and -.31-), and the ITSEA Competence scale with all CBCL2/3 domains (between -.28 and -.31), PSI scales (between -.16 and -.24), and with CCTI Soothability (.28).

Method of Scoring: Web-based scoring is available. For more information, email: Alice.Carter@umb.edu or ITSEA@yale.edu, or call 203-764-9093. Prior to scoring, 7 reverse items and “no opportunity” responses are recoded. Then, the items in the scale are summed and divided by the total number of non-missing items for a given subject. The following scores are calculated: Domain Scores, Scale Scores, and Indices. Tables are provided to derive the T-Scores for the domains and Cut-Points for the domain, scales, and indices. Cut-points are the mean scores for the domains and scales above and below which the extreme 10 percent at either ends of the reference sample fell.

Interpretability: The manual contains tables of the reference group’s mean scores by age and sex group for comparison to the calculated mean score. The manual contains tables for converting raw scores into T-scores and the cut-points for each age group and sex grouping. Cut-points are the mean score for the domains and scales above

and below which the extreme 10 percent at either ends of the reference sample fell. The interpretation of the ITSEA results should be done by a professional trained in the administration and interpretation of psychometric tests.

When following up with parents and interpreting the results, it is important to assess whether the behaviors that the parent endorsed are (1) of concern to the parent; (2) evident in multiple settings and with multiple people; (3) culturally appropriate or inappropriate; (4) atypical or deviant problems in development; (5) transient problems that are secondary responses to stress; (6) transient problems that are related to a child’s developmental phase.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Not described, however, the manual stresses that the ITSEA was not designed to be used or interpreted by parents or other caregivers without the help of a professional who has received training in the administration and interpretation of psychometric tests.

References:

Carter, Alice S., and Margaret J. Briggs-Gowan. *Infant Toddler Social and Emotional Assessment (ITSEA) Manual Version 1.1*. June 13, 2001.

INFANT/TODDLER SYMPTOM CHECKLIST (ITSC), 1995

Authors:

Georgia DeGangi, Susan Poisson, Ruth Sickel, and Andrea Santman Wiener

Publisher:

Therapy Skill Builders, a division of the Psychological Corporation
800-872-1726

Initial Material Cost:

Complete set (Includes manual, 6 sets of 5 score sheets in 25 page pads and vinyl storage portfolio) \$63

Representativeness of Norming Sample: No norming sample.

Languages:

English

Type of Assessment:

Parent report or interview

Age Range and Administration Interval:

7-30 months old

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 10 minutes, and it can be administered by the parent or by a paraprofessional. Special training is not required for administering the instrument; however, an understanding of the domains is critical for an accurate interpretation of findings. Scoring can be done in less than 10 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 2 (less than .5 for concurrent)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered or administered and scored by someone with basic clerical skills)

Description: The Infant/Toddler Symptom Checklist (ITSC) is designed to screen 7 to 30-month-old infants and toddlers for sensory and regulatory disorders who are behaviorally problematic and show disturbances in sleep, feeding, state control, self-calming, and mood regulation. The checklist focuses on infant behavior in the following domains: (1) self-regulation, (2) attention, (3) sleep, (4) eating or feeding, (5) dressing, bathing, and touch, (6) movement, (7) listening

and language, (8) looking and sight, and (9) attachment/emotional functioning. Questions are answered with a “never or sometimes,” “most times,” or “past.” The criterion-referenced ITSC checklist comes in six versions: a single short version for general screening purposes and five age-specific screens for both diagnostic and screening purposes: 7 to 9 months, 10 to 12 months, 13 to 18 months, 19 to 24 months, and 25 to 30 months. The authors recommend using other

observation tools when using ITSC as a screening tool and traditional developmental tests when using it as a diagnostic tool.

Uses of Information: The ITSC is used to determine whether a child may have a predisposition toward developing sensory integrative disorders, attention deficits, or emotional, behavioral, or learning difficulties, and whether further diagnosis is required.

Reliability: None described.

Validity: The research sample consisted of 154 normal and 67 regulatory-disordered infants between 7 and 30 months who were primarily white and middle class. (1) Concurrent validity: Statistical tests were performed on correlations between scores on the ITSC and the Bayley Scales of Infant Development, Mental Scale; the Test of Sensory Functions in Infants (TSFI); and the Test of Attention in Infants (TAI) for a sample of normal infants and a sample of regulatory disorder infants. The results showed that correlations were statistically significant for the regulatory disorder infants and only a few of the correlations with the TSF and TIA subtests were significant for the normal infants, especially among the 7- to 9-month-olds. The authors concluded that the ITSC provided information that is distinct from that obtained by diagnostic measures, particularly for 10- to 30-month olds. The authors tested for construct validity to select the instrument items by performing t-tests on the difference between means obtained from a sample of regulatory disorder infants and a sample of normal infants. They also performed another test by comparing

the scores of parents who did not express concern over their infant's development with those of parents who did express concern and found that only one (out of 25) of the scores in the no concern group was above the at-risk cutoff score while all but two (out of 14) in the concern group had scores about the cutoff. The authors then performed diagnostic tests on infants in the concern group using the TSFI and the TIA and concluded that all 14 suffered from regulatory disorders. (2) Predictive validity: In a separate publication, the authors reported that 78 percent of infants identified by the ITSC as having problems were diagnosed with developmental or behavioral problems at 3 years of age using standardized measures such as the Child Behavior Checklist.

Method of Scoring: The item responses are scored in the following manner: 0 points for "never or sometimes," 1 point for "past," and 2 points for "most times" (a self-calming item is the only item that is scored differently due to different response categories). The points are then summed for the entire checklist. The total score is then compared to a normal score range for the appropriate age group. A protocol sheet is available to assist in the scoring of the instrument.

Interpretability: The total checklist score is compared to the cutoff score for normal functioning infants and toddlers in the child's age group. Children whose scores fall at or above the cutoff score are considered to be at risk of having a regulatory disorder and further diagnosis is warranted.

Training Support: The manual includes case studies.

**Adaptations/Special Instructions for
Individuals with Disabilities:** None described.

Report Preparation Support: None
described.

References:

DeGangi, Georgia A., Susan Poisson, Ruth Z.
Sickel, and Andrea Santman Wiener.
Infant/Toddler Symptom Checklist: A Screening
Tool for Parents. San Antonio, TX: Therapy Skill
Builders, Psychological Corporation, 1995.

LEITER INTERNATIONAL PERFORMANCE SCALE-REVISED (LEITER-R), 1997

Authors:

Gale H. Roid and Lucy J. Miller

Publisher:

Stoelting Co.

(630) 860-9700

www.stoeltingco.com

Initial Material Cost:

Complete Leiter-R Kit: \$850 (includes manual, 3 easel books, response cards, manipulatives, record forms, booklets, and carrying case)

Representativeness of Norming Sample:

The Leiter-R was standardized on 1,719 typical children and adolescents and 692 atypical children ages 2 years to 20 years, 11 months using a national stratification plan based on 1993 U.S. Census statistics for age, gender, and socioeconomic status. Nationally representative proportions of children who are Caucasian, Hispanic-American, African-American, Asian-American, and Native American were included.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval: 2 years to 20 years, 11 months

Personnel, Training, Administration, and Scoring Requirements:

The Leiter-R should be administered by a trained individual who has received supervised training and practice. It should be interpreted by someone with graduate training in psychological assessment. Administration times for the Leiter-R range from 25 to 40 minutes.

Summary

Initial Material Cost: 3 (\$200 or higher)

Reliability: 3 (.65 or higher)

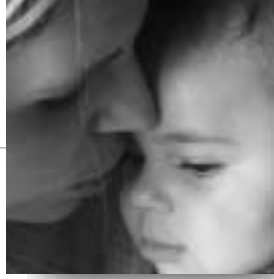
Validity: 3 (concurrent .5 or higher, no information or predictive)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The Leiter-R is an individually administered nonverbal test designed to assess cognitive functions in children and adolescents. It was developed to provide a reliable and valid nonverbal measure of intellectual ability, memory, and attention that could be used to assess children, adolescents, and young adults who could not be reliably and validly tested using traditional

intelligence tests. The Leiter-R consists of two groupings of subtests: (1) the Visualization and Reasoning (VR) Battery (10 subtests), and (2) the Attention and Memory (AM) Battery (10 subtests). It also includes four social-emotional rating scales (Examiner, Parent, Self, and Teacher) that provide behavioral observation information about the examinee. The majority of Leiter-R



items require the child to move response cards into slots on the easel tray. Other items require arranging manipulatives (foam rubber shapes) and pointing to responses on the easel pictures. Starting points in the subtests are determined by the child's age (there are three age groups for administration of the Leiter-R: 2-5, 6-10, and 11-20).

Uses of Information: The Leiter-R can be used to help identify children with cognitive disabilities, to monitor small increments of improvements in cognitive abilities, and to develop intervention strategies that address the identified disabilities. For initial screening purposes, four subtests in the VR Battery can be used to measure the child's global intellectual level as part of a battery of other tests and assessments. The full VR Battery (six subtests for children ages 2 to 5) can be used for identification, classification, and placement decisions. The AM Battery can be used for a comprehensive diagnostic assessment of attention and memory difficulties, neuropsychological evaluations, and evaluation of cognitive process deficits in learning disabilities or attention deficit disorders. Examiners have the option of using the VR and AM Batteries separately; however, the Batteries should be used together for a thorough cognitive assessment, particularly when it is expected that cognitive-process deficits in memory or attention are interfering with the

accurate evaluation of global intellectual level.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): for children age 2, alphas ranged from .71 to .94 across the VR Battery subtests and, for children ages 2 to 3, the alphas ranged from .77 to .89 for the AM Battery subtests. For children ages 2 to 5, the alphas ranged from .71 to .90 for the AM Battery special diagnostic scale, and the reliability coefficients from .94 to .99 for the composite rating scale, and .87 to .93 for IQ and composite scores. (2) Test-retest reliability (with interval not reported): test-retest correlations ranged from .61 to .95 across the VR Battery subtest and composite scores and .86 to .94 across the examiner rating scales and composite for children ages 2 to 5.¹ (3) Inter-rater reliability: no information reported in the manual.

Validity: (1) Concurrent validity: concurrent validity tests between the Leiter-R (Brief and Full Scale IQ) and the Wechsler Intelligence Scale for Children (WISC-III) (Performance and Full Scale IQ) on children ages 6 to 16 resulted in correlations of .85 and .86. The reported tests between Leiter-R Full Scale IQ scores with other cognitive tests showed correlations that ranged from .38 to .66. Tests for accuracy on children ages 2 to 20 showed that a cut-point of 70 on the Leiter-R Full Scale IQ score correctly classified more than 80 percent of children with cognitive delays. The

¹No AM Battery retests were performed on children under the age of 6.

classification accuracy of the Leiter-R for identifying giftedness was not as good, and the manual recommends that the Leiter-R never be used in isolation to identify giftedness. (2) Predictive validity: no information available.

Method of Scoring: The manual contains detailed scoring instructions. For most subtests, responses are scored as (0) Fail or incorrect, or (1) Pass or correct. Scoring criteria for each item are noted on the instruction page for each subtest. For some subtests, scoring requires counting the number of correct responses and the number of errors. Raw scores are typically obtained by summing correct responses. The raw scores on the subtests and rating scales are converted to scaled scores (with a mean of 10 and a standard deviation of 3) using a table provided in the manual. IQ scores are calculated from sums of subtest scaled scores and converted to IQ standard scores (with a mean of 100 and standard deviation of 15) using a table in the manual. Composite scores can also be obtained for Fluid Reasoning, Fundamental Visualization, Spatial Visualization, Attention, and Memory. In addition, the raw scores for each subtest and IQ can be converted to growth-scale scores that define a child's domain of abilities in a metric that can reflect growth and be useful for treatment planning and measuring change over time.

Interpretability: Only persons with graduate training in psychological testing and statistics should interpret the results of the Leiter-R. The manual also cautions that IQ scores from the

Leiter-R should never be used in isolation and should be evaluated in the context of a wide variety of information about the child. The manual includes an extensive discussion of the interpretation of Leiter-R results and provides case studies to demonstrate the interpretation of scores.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: The Leiter-R was specially developed to be used with children who could not be reliably and validly tested using traditional intelligence tests, including children with significant communication disorders, cognitive delays, English as a second language, hearing impairments, motor impairments, traumatic brain injury, attention-deficit disorder, and certain types of learning disabilities. The manual discusses adaptations to administration methods that may be needed for some children to establish that the child understands the nature of the task or to enable the child to communicate answers to test items within the capabilities they have. When such adaptations are made, growth-scale scores that do not rely on normative comparisons should be used.

Report Preparation Support: None described.

References:

Roid, Gale H., and Lucy J. Miller. *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting Co., 1995, 1997.

MACARTHUR COMMUNICATIVE DEVELOPMENT INVENTORIES, 1993

Authors:

Larry Fenson, Philip S. Dale, J. Steven Reznick, Donna Thal, Elizabeth Bates, Jeffrey P. Hartung, Steven Pethick, and Judy S. Reilly

Publisher:

Singular Publishing Group, Inc.
(800) 354-9706
www.singpub.com

Initial Material Cost:

MacArthur Communicative Development Inventories: User Guide, Technical Manual, Words and Gestures, Words and Sentences: \$213

Representativeness of Norming Sample:

Neither the long form nor the short form samples were nationally representative. The long form norming sample consisted of 1,789, and the short form sample included 1,379 children without disabilities living in New Haven, Seattle, and San Diego. Parents were more educated and less ethnically diverse than the general population.

Languages:

English, Spanish, Italian

Type of Assessment:

Parent report

Age Range and Administration Interval: 8 to 30 months

Personnel, Training, Administration, and Scoring Requirements:

No training required to complete forms. Takes between 20 to 40 minutes for a parent to complete and about 10 minutes for a staff member to score an inventory.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (most .65 or higher)

Validity: 3 (most .5 or higher for concurrent)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 2 (self administered and scored by staff)

Description: The MacArthur Communicative Development Inventories (CDI) assess early language skills of children between the ages of 8 and 30 months using parent reports. The CDI consists of two inventories, each with two sections. The CDI/Words and Gestures inventory is for infants between the ages of 8 and

16 months. The inventory's words section, which has a 28-item list of phrases and a 396-word checklist, is used to assess the infant's production and understanding of words and phrases. The gestures section covers 63 gestures for communication, play, imitation of parents and other adults, and activities with objects. The CDI/Words and

Sentences inventory is for toddlers between the ages of 16 and 30 months. The inventory's word section assesses vocabulary using a 680-word checklist. The second part assesses the toddler's use of possessives, plurals, and tenses, and development of complex sentences. A short version of the inventories is also available. The short version of the CDI/Words and Gestures inventory consists of an 89-word vocabulary checklist and there are two equivalent short versions of the CDI/Words and Sentence Inventory, both with a 100-word vocabulary checklist.

Uses of Information: The CDI can be used to screen for delays in language development, to identify problematic skills, help formulate intervention strategies, and evaluate treatment outcomes.

Reliability: For the long form, (1) internal consistency reliability: Cronbach alpha coefficients for the CDI/Words and Gestures vocabulary production, vocabulary comprehension, and gestures scales were .95, .96, and .39, respectively. The low alpha for the gestures scale resulted from the low correlation of two gestures subscales with a third. The CDI/Words and Sentences vocabulary production and sentence complexity scales had alphas of .96 and .95, respectively. (2) Test-retest reliability (6-week and 6-month intervals): correlation coefficients ranged from .60 to .90 on the infant inventory and .90 and higher on the toddler inventory. (3) Inter-rater reliability: no information available.

For the short form, internal consistency: the infant form had a Cronbach alpha of .97 and the

toddler form A and B each had an alpha of .99. Test-retest (with 2-week interval): the infant form had a correlation of .88 for vocabulary comprehension and .90 for vocabulary production. The vocabulary production test-retest correlations were .74 and .93 for the toddler form A and B, respectively.

Validity: For the long form, (1) Concurrent validity: several comparisons of parent reports using the earlier version of the CDI/Word and Sentences inventory (little changed from the current inventory) with assessments made through laboratory observations found correlations that ranged from .40 to .67 when compared to the Preschool Language Scale and .53 to .85 when compared to the Expressive One Word Picture Vocabulary Test. Similar tests on an even earlier version of the inventory with Bayley language subscales yielded correlations ranging from .33 to .79. Tests of the gestures scale showed significant differences between high- and low-gesture children on three of the four laboratory assessment instruments and "substantial and significant concurrent correlations" between parent and laboratory gestures reports for 12-month-old children and one of the gestures subscales at 10 months. Correlations between the CDI/Word and Sentence syntactic development scale and laboratory measures ranged from .74 to .88. (2) Predictive validity: administrations six months apart of the CDI/Words and Sentences inventory for total vocabulary and grammatical complexity yielded correlations of .71 and .62, respectively. The CDI/Words and Gestures inventories administered

six months apart had a total vocabulary correlation of .38. The correlation between the inventories on vocabulary production was .69.

For the short form, the correlations between the infant short and long form were .98 on vocabulary comprehension and .97 on vocabulary production. The overall correlations between the toddler long form and short form A and B were each .99.

Method of Scoring: Scoring can be done manually or by computer. The User's Guide provides instructions for manual scoring. Scoring the inventories usually involves counting the number of marked items or affirmative responses by sections. Using tables in the User's Guide, raw scores can be converted into gender- and age-specific percentile rankings. Scoring software is available at no cost at www.utdallas.edu/~vamarch/cdi/. The software scores both the long and short and English and Spanish versions of the inventories.

Interpretability: The manual provides instructions for interpreting the results. The normed percentile ranking allows the infant/toddler's performance to be compared to other infants/toddlers.

Training Support: None described.

Adaptations/Special Instructions for

Individuals with Disabilities: The manual cautions against using the CDI with developmentally delayed children whose chronological age exceeds the upper limits of the inventory.

Report Preparation Support: The manual contains report forms for each inventory for the user to complete.

References:

Fenson, Larry, Philip S. Dale, J. Steven Reznick, Donna Thal, Elizabeth Bates, Jeffery P. Hartung, Steve Pethick, and Judy S. Reilly. MacArthur Communicative Development Inventories, User's Guide and Technical Manual. Singular Publishing Group, Inc., San Diego, CA, 1993.

Fenson, Larry, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal, and Stephen J. Pethick. Variability in Early Communicative Development. Monographs of the Society for Research in Child Development, Vol. 59, No. 5, 1994.

Fenson, Larry, Steve Pethick, Connie Renda, Philip S. Dale, and J. Steven Reznick. "Short-form Versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics*. Vol. 21, No. 1, 2000, pp. 91-115.

MULLEN SCALES OF EARLY LEARNING, AGS EDITION, 1995

Authors:

Ellen M. Mullen

Publisher:

American Guidance Service, Inc.

(800) 328-2560

(800) 471-7220 (fax)

Customerservice@agsnet.com

Initial Material Cost:

Mullen Kit (birth to 68 months): \$636

Mullen Scales, Upgrade for Infant Mullen (birth to 39 months): \$282

Representativeness of Norming Sample: Nationally representative sample of 1,849 children ages 2 days to 69 months with no known physical and mental disabilities and parents who spoke primarily English. Data on children in the northeast region was collected from 1981 to 1986 and from 1987 to 1989 for the south, west, north, and north central regions.

Languages:

English

Type of Assessment:

Direct child assessment (with some direct parent assessment)

Age Range and Administration Interval: 0 to 68 months

Personnel, Training, Administration, and Scoring Requirements:

For professionals with training or practical experience in the clinical assessment of infants and young children. The scales can be administered in approximately 15 minutes for 1-year-olds and 30 minutes for 3-year-olds.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher concurrent validity)

Norming Sample Characteristics: 3 (nationally representative and mostly within the past 15 years)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Mullen Scales assess the cognitive functioning of young children from birth to 68 months. The assessment is based on the child's responses to activities prepared by the examiner. Believing that a global intellectual performance measure may mask uneven cognitive development, the Mullen Scales measure five distinct skills, Gross Motor and four "cognitive" skills—Fine Motor, Visual Reception, Receptive

Language, and Expressive Language. The gross motor scale is administered to children from birth to 33 months and the four "cognitive" scales are administered to children from birth to 68 months. The "cognitive" scores can be summarized into an Early Learning Composite (ELC) score.

Uses of Information: Mullen scale scores can be used to identify children with special edu-

cation needs who are eligible for further evaluation. The normative scores can also provide an objective means to identify weaknesses and strengths that underlie a child's learning style for the purpose of designing individualized instructional plans that capitalize on the child's strengths.

Reliability: (1) Internal consistency reliability: the median internal consistency split-half coefficients (Guilford's formula) for the five Mullen scales range from .75 to .83 and for the composite, .91. (2) Test-retest reliability (with a 1- to 2-week interval between tests): for the Gross Motor Scale of the original Mullen scales, the correlation between tests was .96, and the median correlations on the "cognitive" scale were .84 (with a range of .82 to .85) for children ages 1 to 24 months and .76 (with a range of .71 to .79) for children ages 25 to 56 months. (3) Inter-rater reliability: correlations among raters ranged from .91 to .99 for age groups between 1 and 44 months.

Validity: (1) Concurrent validity: tests showed the Mullen scales to have stronger correlations with instruments that measured similar skills than those measuring different skills. The correlations of the Mullen "cognitive" scales with the Bayley Mental Development Index (MDI) were higher (.53 to .59) than their correlations with the Bayley Psychomotor Development Index (PDI; .21 to .52). The ELC also was more strongly correlated with the MDI (.70) than with the PDI (.43). Conversely, the Mullen Gross Motor scale was more strongly correlated with the Bayley PDI (.76) than with the MDI (.30). Similarly, the

Mullen Receptive Language scale had a higher correlation with the Preschool Language Assessment Auditory Comprehension (.85) than with Verbal Ability (.72), while the converse was true with the Mullen Expressive Language Scale (.72 for auditory and .80 for verbal). Finally, the Mullen Fine Motor scale was strongly correlated with the Peabody Fine Motor Scale, across four age groups of children between the ages of 6 and 36 months (correlations ranged from .65 to .82). (2) Predictive validity: no information available.

Method of Scoring: The Item Administration Book provides instructions for scoring the items on the scales. Scoring is done on a record form containing a list of tasks or stimuli of possible responses for each assessment item. In most cases, the child receives a "1" for correct responses and "0" for incorrect responses. In some cases, the tester must sum the task scores to obtain the item scores. There are also cases where the item score can range from anywhere between 0 and 5. Scoring software (ASSIST) is available for purchase.

Interpretability: The raw scores for each scale can be converted into age-adjusted normalized scores. The four "cognitive" skills T score can be further converted into a normalized ELC score, which has a mean of 100 and a standard deviation of 15. In addition, the scores can be used to obtain the child's percentile rank and age equivalent score, the age at which the child's raw score is the median score. The manual provides instructions for interpreting these scores, taking into account variables that may influence them. The ASSIST

software program converts raw scores into the normalized scores and provides interpretative information.

Training Support: A training videotape can be purchased for \$104.95.

Adaptations/Special Instructions for Individuals with Disabilities: None.

Report Preparation Support: The manual

provides three case studies as examples of how the Mullen Scales can be used and reported. The Mullen ASSIST computer software program provides an optional narrative report.

References:

Mullen, Eileen M. Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Services, Inc., 1995.

THE OUNCE SCALE, 2002 ¹

Authors:

Samuel J. Meisels, Amy Dombro, Dot Marsden, Donna Weston, and Abby Jewkes

Publisher:

Pearson Early Learning
(800) 552-2259
www.pearsonearlylearning.com/index.html

Initial Material Cost:

Not yet available as of July 2002

Representativeness of Norming Sample:

Not yet available as of July 2002

Languages:

English (Spanish version of the Family Album is being developed)

Type of Assessment:

Observation, parent report.

Age Range and Administration Interval:

Eight intervals from 0 to 42 months:

0 to 4 months 18 to 24 months

4 to 8 months 24 to 30 months

8 to 12 months 30 to 36 months

12 to 16 months 36 to 42 months

Personnel, Training, Administration, and Scoring Requirements:

Not available

Summary Not yet available as of July 2002

Initial Material Cost: Not available

Reliability: Not available

Validity: Not available

Norming Sample Characteristics: Not available

Ease of Administration and Scoring: Not available

Description: The Ounce Scale is an observational assessment for evaluating infants' and toddlers' development from birth to 3-1/2 years. The Ounce Scale has three elements: (1) the Observation Record, which provides a focus for observing and documenting children's everyday behaviors and provides data for making evaluations about development, (2) the Family Album, which provides a structure for parents to learn about and record their child's development as they write down what they see, using photos,

telling stories, and responding to observation questions that are the same as the ones in the caregiver's Observation Record, and (3) the Developmental Profile, which enables caregivers and other staff to evaluate each child's development and progress over time, comparing their observation data to specific performance standards. The Ounce Scale is organized around six major areas of development: Personal Connections, Feelings About Self, Relationships With Other Children, Understanding and

¹ The Ounce Scale, or the Ounce of Prevention Scale, represents work in progress, expected for release in Fall 2002.

Communicating, Exploration and Problem Solving, and Movement and Coordination.

Uses of Information: The Ounce Scale's purpose is twofold: (1) to provide guidelines and standards for observing and interpreting young children's growth and behavior, and (2) to provide information that parents and caregivers can use to plan curriculum and engage in enhancing activities and experiences with infants and toddlers.

Reliability: Not yet available as of July 2002

Validity: Not yet available as of July 2002

Method of Scoring: The service provider summarizes the Observational Record and the Family Album in the Developmental Profile. Behaviors are marked as "Developing as Expected" or "Needs Development" and space is available to record comments.

Interpretability: Not yet available as of July 2002

Training Support: Not yet available as of July 2002

Adaptations/Special Instructions for Individuals with Disabilities: Not yet available as of July 2002

Report Preparation Support: The Family

Album helps parents keep track of their child's growth and development, assisted or mentored by service providers. Also, the Developmental Profile, which is a summary and evaluation of the child's performance and progress over four- to six-month intervals, is summarized by the service provider and shared with the parent or caregiver. In the Developmental Profile conferences, supervisors review the information and documentation that the caregivers have about each child and family, go over the Observational Record, and discuss the rationale for the evaluations made on the Developmental Profile. The supervisors also discuss how the caregiver will share the Profile with the parents and outline future steps the caregiver is planning to advance or assist the child's and family's growth and development.

References:

Meisels, Samuel J. "Fusing Assessment and Intervention: Changing Parents' and Providers' Views of Young Children." *Zero to Three*, vol. 21, no. 4, February/March 2001, pp. 4-10.

Meisels, Samuel J., Amy Laura Dombro, Dorothea B. Marsden, Donna R. Weston, and Abigail M. Jewkes (in draft). *The Ounce Scale*. Ann Arbor, MI: Rebus

PEABODY PICTURE VOCABULARY TEST, THIRD EDITION (PPVT-III), 1997

Authors:

Lloyd M. Dunn and Leota M. Dunn

Publisher:

American Guidance Services

(800) 328-2560

www.agsnet.com

Initial Material Cost:

PPVT-III Form IIIA and IIIB Test Kit: \$155 for each, or \$263 for both (includes picture plates, 2 norms booklets, 2 examiner's manual, and 1 package (25 performance records))

PPVT-III Form IIIA and IIIB Test Kit with ASSIST: \$254

PPVT-III Form IIIA and IIIB Test Kit with ASSIST: \$362

Representativeness of Norming Sample:

The norming sample consisted of a stratified random sample of 2,725 persons ages 2.5 to over 90 selected to proportionately match the population distribution in the March 1994 Current Population Survey on gender, race/ethnicity, geographic region, and socioeconomic states. The sample was restricted to individuals who understood English. The tests were conducted between March 1995 and March 1996 at 240 nationwide sites.

Languages:

English (see the Test de Vocabulario en Imagenes Peabody, for the Spanish version of the test)

Type of Assessment:

Direct child assessment

Age Range and Administration Interval: -2_ to 90 years.

Personnel, Training, Administration, and Scoring Requirements:

Although formal training in psychometrics is not required, the examiners should be thoroughly familiar with the test materials and well-trained in administering and scoring the test. It takes approximately 12 minutes to administer 5 sets.

Summary

Initial Material Cost: 2 (\$100 to \$200) Reliability: 3 (.65 or higher)

Validity: 1 (validity information not available on children from 0-3)

Norming Sample Characteristics: 3 (normed within the past 15 years with a nationally representative population)

Ease of Administration and Scoring: 2 (administered and scored by a trained individual)

Description: The PPVT-III tests for listening comprehension of standard English. It has two parallel forms, Forms IIIA and IIIB, each with

four training items (for administering the test) and 204 test items grouped into 17 sets of 12 items arranged in order of increasing difficulty.

In the test, the examiner orally presents a stimulus word with a set of pictures and the test taker is asked to select the picture that best represents the word's meaning. The examiner administers the item sets until the child's "basal" and "ceiling" sets are found, a process that usually takes five item sets. The basal set is the item set in which the child makes one or no errors and the ceiling set is the item set in which the child makes eight or more errors. Because it requires no reading or writing and is easy to administer, PPVT-III is useful in testing preschool children and is fair to persons with written-language problems and disabilities such as autism, withdrawn personalities, psychotic symptoms, severe cerebral palsy, and moderate visual disabilities.

Uses of Information: PPVT-III can be used to measure a person's receptive vocabulary. PPVT-III can (1) screen for verbal ability, giftedness, and mental retardation in persons for whom English is the spoken language in the home, community, and school; and (2) measure English language proficiency for persons for whom English is a second language.

Reliability: (1) Internal consistency validity (Cronbach's alpha): .93 for Form IIIA for both ages 2 years 6 months and 3 years. For Form IIIB, the alphas were .93 and .92, respectively. The Spearman-Brown split-half reliability coefficients for Form IIIA were .89 for ages 2 years, 6 months and .91 for age 3, and for Form IIIB .92 and .93, respectively. (2) Test-retest reliability (with an approximately one-month interval between tests): the correlation coefficients for ages 2 years, 6

months through 5 years, 11 months were .92 for Form IIIA and .91 for Form IIIB.

Validity: (1) Concurrent validity: Tests that compared PPVT-III with intelligence tests were conducted and the results were corrected for norm group variability. Corrected correlation coefficients ranged from .82 to .92 with the Wechsler Intelligence Scale for Children—Third Edition verbal, performance, and full scale IQ scales, .76 to .91 with the Kaufman Adolescent and Adult Intelligence Test crystallized, fluid, and composite IQ tests, .62 to .82 with the Kaufman Brief Intelligence Test vocabulary, matrices, and composite tests, and .63 to .83 with the Oral and Written Language Scales listening comprehension, oral expression, and composite scales. However, all of these tests were performed on persons older than 3. (2) Predictive validity: no information available.

Method of Scoring: The examiner enters the number associated with the picture selected by the examinee on the performance record form and draws a "slash" through the letter "E" under the error column for each error. The raw score is obtained by subtracting the total number of errors above the examinee's basal set from the number associated with the most difficult item answered correctly in the examinee's ceiling set. Using several tables, raw scores can be converted into different types of age-adjusted standardized scores. One table converts the raw score into a standard score. A second table converts the standard score into percentile rank, normal curve equivalent, and stanine. Finally, a third table con-

verts the examinee's raw score into the age equivalent performance. A scoring software program is available for purchase. The software scores the test and converts the raw scores into standard scores, percentiles, stanines, normal curve equivalents, and age equivalents.

Interpretability: Only persons with formal training in psychological testing and statistics should interpret the results of the PPVT-III.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: Because no reading or writing is required, the PPVT-III can be administered to many groups with disabilities without any changes. The examiners' manual describes various modifications that can be made in administering the test to accommodate various disabilities.

Report Preparation Support: None described.

References:

American Guidance Service. PPVT-III ASSIST. Circle Pine, MN: American Guidance Service, 1997.

Dunn, Lloyd M. and Leota M. Dunn. Examiner's Manual for the Peabody Picture Vocabulary Test—Third Edition. Circle Pines, MN: American Guidance Service, Inc., 1997.

Dunn, Lloyd M. and Leota M. Dunn. Norms Booklet for the Peabody Picture Vocabulary Test—Third Edition. Circle Pines, MN: American Guidance Service, Inc., 1997.

Williams, Kathleen T. and Jing-Jen Wang. Technical References to the Peabody Picture Vocabulary Test—Third Edition. Circle Pines, MN: American Guidance Service, Inc., 1997.

PRESCHOOL LANGUAGE SCALE (PLS-3), 1992 ¹

Authors:

Irla Lee Zimmerman, Violette G. Steiner, Roberta Evatt Pond

Publisher:

The Psychological Corporation
(800) 872-1726
www.psychcorp.com

Initial Material Cost:

\$185 (includes Examiner's Manual, Picture Manual, and 12 Record Forms)

Representativeness of Norming Sample:

Sample was based on the 1980 Census of Population (1986 update) stratified on the basis of parent education level, geographic region, and race/ethnicity. The standardization study, which began after June 1991, excluded children with language disorders, children who were more than a month premature, and children who experienced difficulties at birth.

Languages:

English and Spanish

Type of Assessment:

Direct assessment

Age Range and Administration Interval: 2 weeks to 6 years

Personnel, Training, Administration, and Scoring Requirements:

Must have knowledge of test administration and score interpretation as well as know the instrument and adhere to administration procedures. Requires 15 to 30 minutes to administer the PLS-3 for children under 1 year and 30 to 40 minutes for children 12 months to 3 years, 11 months old.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 1 (validity not described for 0-3 age group)

Norming Sample Characteristics: 3 (Nationally representative, less than 15 years old)

Ease of Administration and Scoring: 3 (administered and scored by highly trained individual)

Description: The Preschool Language Scale (PLS-3) provides clinicians with a diagnostic instrument to evaluate language development. The PLS-3 can be used to test receptive and expressive language skills with children from 2

weeks through 6 years of age. It also assesses behaviors considered to be language precursors. The PLS-3 contains two standardized subscales and three supplemental measures (these measures are not incorporated into the test scores, but they

¹ PSL-4 will be available in summer 2002.

are optional and provide additional information). The standardized subscales are Auditory Comprehension (focusing on attention abilities) and Expressive Communication (focusing on social communication and vocal development). The supplemental measures include the Articulation Screener, Language Sample Checklist (LSC), and Family Information and Suggestions Form (FISF). The LSC evaluates the child's language skills in conversational speech. The FISF, which is completed by the parents, is for information on the child's communication behavior at home and the needs of the family and areas that parents would like to see addressed. The Preschool Language Scale includes eight receptive/expressive language tasks for each six-month interval for ages birth through 4 years, 11 months; and eight receptive/expressive tasks for each 12-month interval for ages 5 and 6 years.

Uses of Information: The PLS-3 can be used to determine whether or not a child has a language disorder and, if so, whether the source of the disorder is auditory, expressive, or an overall problem. The standard score and percentile ranks obtained can also help determine the severity of the disorder and identify areas for in-depth testing before defining therapy goals.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): for children ages birth to 2 years, 11 months, alphas ranged from .47 to .86 for Auditory Comprehension, from .68 to .86 for Expressive Communication, and from .74 to .92 for the Total Language Score. (2) Test-retest reliability: no information available. (3) Inter-rater

reliability: .98 (based on observations of children older than age 3).

Validity: (1) Concurrent validity: For children older than 3, the correlations with the Clinical Evaluation of Language Fundamentals-Revised (CELF-R) were .69 for Auditory Comprehension, .75 for Expressive Communication, and .82 for Total Language. Results of other concurrent validity studies can be obtained by the publisher. (2) Predictive validity: no information available.

Method of Scoring: In most cases, response items are scored by entering a check for each correct response, a minus for each incorrect response, and an "NR" if the child does not respond. A few items are open-ended, requiring a written response. A score graph in the Record Form enables the clinician to visually compare a child's PLS-3 Auditory Comprehension and Expressive Communication scores. PLS-3 provides age-based standard scores, percentile ranks, and age equivalents for the Auditory Comprehension, Expressive Communication, and Total Language scores.

Interpretability: The Examiner's Guide provides detailed information on interpretation of scores. The PLS-3 includes scoring interpretation guidelines as they relate to determining the severity of the disorder and the need for intervention. Using the task analyses (the PLS-3 Checklist and Profile), a clinician can evaluate the child's strengths, emerging skills, and deficits. The Checklist groups PLS-3 tasks by age; the Profile groups tasks by the type of language skill tested.

Training Support: None mentioned in manual

Adaptations/Special Instructions for Individuals with Disabilities: The examiner's manual includes instructions for administering the PLS-3 to individuals who appear to have severe developmental delays, severe physical impairments, or hearing impairments.

Report Preparation Support: The Record Form includes a Clinician's Worksheet to help summarize information from the assessment and from the FISF and to plan for follow-up testing and treatment. The worksheet has three sections: assessment results, follow up for child and family, and outcomes of additional assessment and training. The worksheet is designed to include infor-

mation relevant to the development of the Individualized Family Service Plan.

References:

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. *Preschool Language Scale - 3: Examiner's Manual*. San Antonio, TX: The Psychological Corporation, 1992.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. *Preschool Language Scale - 3: Picture Manual*. San Antonio, TX: The Psychological Corporation, 1992.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. *Preschool Language Scale - 3: Record Form*. San Antonio, TX: The Psychological Corporation, 1992.

RECEPTIVE-EXPRESSIVE EMERGENT LANGUAGE TEST - SECOND EDITION (REEL-2), 1991

Authors:

Kenneth R. Bzoch and Richard League

Publisher:

PRO-ED

(800) 897-3202

www.proedinc.com

Initial Material Cost:

REEL-2 Complete Kit: \$86 (includes Examiner's Manual and 25 Profile/Test Forms)

Representativeness of Norming Sample:

Not representative, norming sample included only white infants.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval: 0 to 3 years**Personnel, Training, Administration, and Scoring Requirements:**

It is possible for a consultant or expert with clinical training to learn to administer the REEL-2 in two hours. The manual suggests that a trainee be observed by an individual with REEL-2 experience.

Administration time per child is 10 to 15 minutes; scoring time is 5 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher)

Validity: 1 (not available)

Norming Sample Characteristics: 2 (only white infants included)

Ease of Administration and Scoring: 3 (administered and scored by a clinician or expert with clinical training)

Description: The REEL-2 assesses communication behaviors (receptive and expressive) of infants and toddlers from birth to 3 years of age. The format is a 132-item checklist of language milestones, with 3 items contained in each of the 22 age intervals (age intervals vary from 1 to 3 months depending on the chronological age of the child). The examiner completes the checklist based on information provided by the child's caregiver. Typically, more than three items need

to be administered to obtain the ceiling age interval for the child (further details are provided in the "Method of Scoring" section below). The REEL-2 uses caregiver report to identify any major language problems. Scores derived from the REEL-2 include an Expressive Language Quotient, a Receptive Language Quotient and a Language Quotient. The REEL-2 is undergoing revision, and the publisher expects the third edition to be released in late 2003 or mid-2004.

Uses of Information: The primary uses of the REEL-2 are to (1) provide descriptions of the developmental status of young children in the language area, (2) assist with setting intervention goals, and (3) serve as a screening instrument for medically and environmentally at-risk populations. The REEL-2 is also used to determine the extent to which interventions have changed the language status of individual children.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): .98 for 0 to 11 months, .97 for 12 to 23 months, and .95 for 24 to 36 months, with an average of .97. (2) Test-retest reliability: .90 to 1.0.

Validity: (1) Concurrent validity: studies showed that the REEL-2 relates well to normal expected levels of functioning (Eich, 1971). However, studies were completed on small samples with similar characteristics. (2) Predictive validity: no information available.

Method of Scoring: The examiner uses the child's chronological age to determine the age interval from which the questions should first be asked. Behaviors observed for each item are scored as either typical of the child (+), emergent (+-), or never observed (-). Scores are computed by summing the (+) responses. The examiner then determines the ceiling interval, which is the highest age interval receiving at least two (+) item scores. The results of the Receptive Language Age and the Expressive Language Age are combined to form the Combined Language Age. Then, each of these scores (the Receptive Language Age, Expressive Language Age, and the Combined

Language Age) are respectively divided by the child's chronological age and multiplied by 100 to obtain a Receptive Quotient, an Expressive Quotient, and a Language Quotient.

Interpretability: The Examiner's Manual includes average scores for the limited norming sample of children used to determine the validity of the REEL-2, but does not contain any overall score percentiles or cutoffs to guide interpretation of the REEL-2 scores. However, it does contain several interpretation guidelines, including the relationship between Receptive Language skills and hearing and nervous system disorders, as well as the relationship between delays in Receptive Language versus Expressive Language. It also includes suggestions for intervention strategies related to language stimulation.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: The Examiner's Manual contains general suggestions on how to present reports/recommendations to parents: (1) the interviewer should provide parents/caregivers with the REEL-2 manual and other relevant booklets; (2) information about the score should be accompanied by the interviewer's explanation, which could include other interpretations and recommendations for intervention programs; and (3) the interviewer should avoid using technical language to explain the REEL-2 scores.

References:

Bzoch, Kenneth, and Richard League.

Receptive-Expressive Emergent Language Test: A Method for Assessing the Language Skills of Infants, Second Edition. Austin: Pro-Ed, 1991.

Bzoch, Kenneth, and Richard League.
Receptive-Expressive Emergent Language Test: A

Method for Assessing the Language Skills of Infants, Second Edition: Examiner's Manual. Austin: Pro-Ed, 1991.

Eich, W.F. A Validation Study for the REEL. (Unpublished manuscript), 1971.

THE ROSSETTI INFANT-TODDLER LANGUAGE SCALE: A MEASURE OF COMMUNICATION AND INTERACTION, 1990

Authors:

Louis Rossetti

Publisher:

LinguiSystems

(800) 776-4332

www.linguisystems.com

Initial Material Cost:

Complete kit: \$70 (includes Examiner's Manual plus 1 package of forms, enough for 10 assessments)

Representativeness of Norming Sample: Not representative. Tool is based on the research and experience of the author.

Languages:

English

Type of Assessment:

Because responses are elicited, observed, and reported, this measure is a combination of direct child assessment, observation, and parent report.

Age Range and Administration Interval:

Birth to 3 years; administered based on child's age in 3-month groupings.

Personnel, Training, Administration, and Scoring Requirements:

Designed for someone with a background in clinical assessment of child development and language. Training requires familiarizing oneself with the manual and questionnaires. Administration of the test will vary depending on whether or not the Parent Questionnaire was filled out and mailed prior to the test day as well as the age of the child (approximately 45 minutes for each age interval). Scoring is done concurrently with the test.

Summary

Initial Material Cost: 1 (<100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a clinician or expert with clinical training)

Description: The Rossetti Infant-Toddler Language Scale assesses the language skills of children from birth through 3 years of age. The scale assesses preverbal and verbal areas of communication and interaction through direct observation

and caregiver report. Areas assessed include (1) Interaction-Attachment (relationship between the caregiver and the infant), (2) Pragmatics (the way language is used to communicate and affect others), (3) Gestures, (4) Play (both individual and

interactive), (5) Language Comprehension, and (6) Language Expression. The test consists of a parent questionnaire and an examiner's evaluation form that address each of the 6 areas. The parent questionnaire allows the examiner to familiarize himself with the developmental concerns expressed by the parent and helps to determine the age level at which testing should begin. The examiner's job is to establish both a baseline and ceiling developmental age by observing, eliciting, or using a caregiver's report of various behaviors listed in each of the six developmental areas.

Uses of Information: The primary use of the Rossetti Infant-Toddler Language Scale is the early detection of language delays in infants and toddlers. When assessment results indicate a child needs early language intervention, a therapy program is developed with specific goals.

Reliability: This measure has not been standardized, and there is no statistical information on it.

Validity: This measure has not been standardized, and there is no statistical information on it.

Method of Scoring: If a parent questionnaire is completed, the description of current skills helps to determine the age level at which testing should begin. Otherwise, the examiner uses the child's chronological age to determine the age interval from which the questions should first be asked. A baseline and ceiling level of performance is established in each of the six developmental areas. To establish a baseline level (all items are mastered in the developmental area), the assess-

ment is begun at six months below the child's chronological age or suspected developmental level. Once a baseline level is established, testing proceeds forward until the child fails all items for a developmental area at a particular age range (ceiling level). Items are considered "passed" if the behavior in question is noted in one of the 3 following ways: (1) Observe (O); (2) Elicit (E); (3) Report (R).

Interpretability: An individual baseline and ceiling age level for each of the six developmental areas may be reported in order to determine the child's relative areas of strengths or weaknesses. In addition, an examiner can compute a global baseline and ceiling age level, which is the oldest age level at which the child mastered all items across all developmental areas. The global basal and ceiling can provide information about the child's performance ability to compare to his/her chronological or adjusted age.

Training Support: The examiner is free to call the publisher with questions related to the administering of the test. In addition, the manual has tips on how to elicit responses or when and where to look for them.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: The Examiner's Manual contains general suggestions on how to present reports and recommendations to parents. The examiner should (1) remain cautious of providing long-term predictions about the child's potential and needs; (2) remain sensitive to the amount of detail that is offered during the initial

conference; (3) schedule a second conference to go over the results in greater detail; (4) actively involve the caregivers in the conference as soon as possible and ask for feedback from the caregiver; and (5) present programming recommendations as options rather than requirements.

References:

Rossetti, Louis. *The Rossetti Infant-Toddler Language Scale: A Measure of Communication and Interaction*. East Moline, IL: LinguiSystems, 1990.

TEMPERAMENT AND ATYPICAL BEHAVIOR SCALE (TABS) – EARLY CHILDHOOD INDICATORS OF DEVELOPMENTAL DYSFUNCTION, 1999

Authors:

John T. Neisworth, Stephen J. Bagnato, John Salvia, and Frances M. Hunt.

Publisher:

Paul H. Brookes Publishing Co.
(800) 638-3775
www.brookespublishing.com

Initial Material Cost:

Complete TABS System: \$85 (includes the manual, a pad of Screeners, and a packet of Assessment Tools)
Manual only: \$40

Representativeness of Norming Sample:

621 children of diverse racial and ethnic backgrounds between ages of 11 and 71 months; 52 percent were 2 years old or younger. Various geographic and socioeconomic conditions. Two separate samples were tested: children with disabilities and those without disabilities. Precisely representative normative samples were not necessary, because the occurrence of atypical behaviors in infants and young children is presumed to be unrelated to various demographic factors.

Languages:

English

Type of Assessment:

Parent report (a professional familiar with child's behavior may also report)

Age Range and Administration Interval:

11 to 71 months

Personnel, Training, Administration, and Scoring Requirements:

Child's parent(s) or a professional familiar with child's behavior can administer the TABS Screener and Assessment Tool. Administration time per child is 5 minutes for the Screener and 15 minutes for the Assessment Tool. Total time for assessment, scoring and interpretation is 30 minutes for each child.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher)

Validity: 1 (none described); content and construct validity are promising; concurrent and predictive validity not mentioned.

Norming Sample Characteristics: 3 (normed within past 15 years, representative of population program works with).

Ease of Administration and Scoring: 2 (administered by parent or professional who is familiar with child's daily behavior)

Description: The TABS assesses atypical temperament and self-regulatory behaviors among infants and young children ages 11 to 71 months that may put them at risk for developmental delay. There are two components to the assessment: (1) a Screener, which is a one-page, 15-item checklist with "Yes" or "No" responses

that is used to determine whether further assessment is needed and (2) an Assessment Tool, which is a 55-item checklist with "Yes," "No" and "Need help" responses, used to obtain a total raw score, or the Temperament and Regulatory Index (TRI). Raw scores are also provided for the four subtests on the following behaviors: detached,

hyper-sensitive/active, underreactive, and dysregulated.

Uses of Information: The primary uses of the TABS are: (1) to serve as a screening tool and determine eligibility for particular services, (2) to identify serious developmentally dysfunctional behaviors early and intervene to reduce their occurrence, (3) to provide assistance to parents in managing atypical behaviors, (4) to plan programs for education, treatment, and intervention, and (5) to evaluate program impacts and conduct research.

Reliability: (1) Internal consistency (Cronbach's alpha) for the TABS Screener is .83. (2) The corrected split-half reliability for the TABS is .95 for children with disabilities and for the pooled samples and .88 for children not at risk.

Validity: Content validity has been ensured in two ways: (1) TABS items have been developed based on an extensive literature review that identified behaviors related to a number of infant and child disorders, and (2) the four factors underlying the TABS have shown promising results when tested statistically. There was no discussion of concurrent or predictive validity.

Method of Scoring: In administering the Screener and the Assessment Tool, the examiner will have checked a "Yes" if a certain behavior is a current or recent problem and a "No" if the behavior is not a problem or does not apply because of the child's age. In scoring the Screener, the examiner adds up the items that are checked "Yes" and that sum becomes the raw score, which

serves as an estimate for the Temperament and Regulatory Index (TRI). Children who have raw scores of 1 or higher on the Screener should be assessed using the Assessment Tool, which more accurately calculates the TRI. There are three types of scores that can be derived from the TABS Assessment Tool: percentiles, standard scores, and normalized standard scores. Typically, the percentiles are the most practical tool for interpreting TABS results for parents, as well as for determining educational and treatment plans for children.

Interpretability: The authors suggest that the child's TRI score should be reported, interpreted and used for decision-making purposes, as the subtest scores tend to be less reliable than the complete TRI. A percentile means that a child has scored equal to or better than the respective percentage of children in the normative sample. For example, a percentile of 70 indicates that a child has a raw score equal to or better than 70 percent of the sample on the TRI. The cutoff point for the TRI is 10. This means that children who score 10 or higher can most likely be classified as having atypical development, meaning they have difficulty with their temperament and self-regulation. Children with disabilities are more likely to earn higher raw scores (6+) while those without disabilities are more likely to earn lower raw scores (0 to 4). A child is at risk for atypical temperament and/or self-regulation if s/he has a TRI score between 5 and 9, because children with disabilities earn those scores equal to or more often than their non-disabled peers.

Training Support: "Brookes on Location"

professional development seminar, Using TABS to Identify Early Atypical Behavior is available through publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described, but the measure was normed with children with and without disabilities.

Report Preparation Support: Chapter 6 describes an early intervention program to address challenging behavior that was developed

using research editions of TABS. A case study is also provided in this chapter to demonstrate how TABS might be used in an early intervention program of this nature.

References:

Neiswroth, John, Stephen Bagnato, John Salvia, and Frances Hunt. TABS Manual for the Temperament and Atypical Behavior Scale: Early Child Indicators of Developmental Dysfunction. Baltimore: Paul H. Brookes Publishing Co., 1999.

TEST DE VOCABULARIO EN IMAGENES PEABODY (TVIP), 1986

Authors:

Lloyd M. Dunn, Eligio R. Padilla, Delia E. Lugo, and Leota M. Dunn

Publisher:

American Guidance Services
(800) 328-2560
www.agsnet.com

Initial Material Cost:

TVIP Test Kit: \$119 (includes test easel, English or Spanish manual, and 25 record forms)
TVIP Manual: \$40
25 Record Forms: \$27

Representativeness of Norming Sample:

The norming sample consisted of monolingual, Spanish-speaking students in Latin America. Testing in Mexico took place between September 1981 and November 1982 and included 1,219 children from the public schools of Mexico with 20% coming from Mexico City. Testing in Puerto Rico took place between September 1982 and February 1983 and included

1,488 children from Puerto Rico with 62.2% from the San Juan metropolitan area. To correct for unevenness of socioeconomic status (SES) representation, a weighting system was used to increase or decrease the contributions of each individual's score at each age, so as to fit the SES ratios established by the U.S. census statistics.

Languages:

Spanish (the Peabody Picture Vocabulary Test is the English Version of this test)

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

2 1/2 to 18 years

Personnel, Training, Administration, and Scoring Requirements:

Although formal training in psychometrics is not required, the examiners should be thoroughly familiar with the test materials and well-trained in administering and scoring the test. It is extremely important that the examiner be proficient in correctly pronouncing each stimulus word. It takes 10 to 15 minutes to administer.

Summary

Initial Material Cost: 2 (\$100 to \$200) Reliability: 3 (.65 or higher for internal consistency, test-retest and inter-rater not described)

Validity: 1 (validity information not available on children from 0-3)

Norming Sample Characteristics: 2 (older than 15 years, not nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The TVIP is based on the Peabody Picture Vocabulary Test-Revised (PPVT-R) and includes 125 translated items to assess the vocabulary of Spanish-speaking and bilingual children. Items were selected through item analysis for their universality and appropriateness to Spanish-speaking communities. In the test, the examiner orally presents a stimulus word with a set of pictures and the test taker is asked to select the picture that best represents the word's meaning. The examiner administers the items until the child's "basal" and "ceiling" are found. The basal is the highest set of eight consecutive correct responses and the ceiling is the lowest set of eight consecutive responses containing six errors. Because it requires no reading or writing and is easy to administer, the TVIP is useful in assessing older toddlers and preschool children and is fair to persons with written-language problems and disabilities such as autism, withdrawn personalities, psychotic symptoms, severe cerebral palsy, and moderate visual disabilities.

Uses of Information: The TVIP can be used to measure a child's receptive or hearing vocabulary of single Spanish words. It may be used as a screening test of verbal ability or verbal intelligence when Spanish is the language of the home and community into which the child was born and when Spanish is, and has been, the primary language of instruction in the child's program. It may also be used as an achievement test showing the extent of Spanish vocabulary acquisition.

Reliability: (1) Internal consistency reliability (split-half reliability): the median correlation

coefficient, corrected using the Spearman-Brown formula, was .93. For age 2_ to 3, the coefficient was .80. (2) Test-retest reliability: no information available. (3) Inter-rater reliability: no information available.

Validity: (1) Concurrent validity: Correlations ranged from .25 to .59 between scores on the TVIP and the Kaufmann-ABC Global Scales and from .28 to .69 between the TVIP and the Kaufman-ABC Achievement Scale Subtests among children ages 3 to 6. The correlation between TVIP and the Habilidad General Ability test was .44 among children attending an urban private school in Puerto Rico. (2) Predictive validity: no information available.

Method of Scoring: The raw score is obtained by subtracting the total number of errors between the basal and ceiling sets from the number associated with the highest item in the ceiling set. Using several tables, raw scores can be converted into different types of age-adjusted standardized scores using Mexican norms, Puerto Rican norms, or norms for a composite group. One table converts the raw score into a standard score. A second table converts the standard score into percentile rank, decile, and stanine. Finally, a third table converts the examinee's raw score into the age equivalent performance.

Interpretability: Only persons with graduate training in psychological testing and statistics who are familiar with the research literature on the language and cognitive development of Hispanic children should interpret the results of the TVIP.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: Because no reading or writing is required, the TVIP can be administered to many groups with disabilities without any changes.

Report Preparation Support: None described.

References:

Dunn, Lloyd M., Eligio R. Padilla, Delia E. Lugo, and Leota M. Dunn. Examiner's Manual for the Test De Vocabulario en Imagenes Peabody: Adaptacion Hispanoamericana. Circle Pines, MN: American Guidance Service, Inc., 1986.

Dunn, Lloyd M., and Leota M. Dunn. Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service, Inc., 1981.

VINELAND ADAPTIVE BEHAVIOR SCALES (VABS), 1984

Authors:

Sara S. Sparrow, David A. Balla, and Dominick V. Cicchetti

Publisher:

American Guidance Service
www.agsnet.com/index.asp
1-800-328-2560

Initial Material Cost:

Complete Vineland Starter Set (Includes Survey Form Starter Set, Expanded Form Starter Set, and Classroom Edition Starter Set): \$170

Representativeness of Norming Sample:

The standardization sample for the Survey Form and Expanded Forms each consists of a nationally representative sample of 3,000 children from birth to 18 years old, stratified by age, race/ethnicity, gender, geographical region, and parental education attainment to reflect the distribution in the 1980 U.S. Census. National non-representative samples of different handicapped persons older than age 5 were used to derive supplementary norms for handicapped individuals.

Languages:

English and Spanish ¹

Type of Assessment:

Individual parent/caregiver interviews

Age Range and Administration Interval:

Survey and Expanded Forms: Birth through 18 years, 11 months, including low-functioning adults

Personnel, Training, Administration, and Scoring Requirements:

Administration time: Survey Form: 20 to 60 minutes; Expanded Form: 60 to 90 minutes; Individual administering the VABS should be a psychologist or equivalent, or social worker

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 1 (.65 or higher)

Validity: 3 (.5 or higher for concurrent)

Norming Sample Characteristics: 2 (older than 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual).

Description: The Vineland Adaptive Behavior Scales (VABS) are designed to assess handicapped and non-handicapped persons in their personal and social functioning. There are two versions of the revised Vineland that can be used with infants and toddlers². Each version differs in the number

¹ Materials available in Spanish include the Survey Form: Record Booklet and Report to Parents and the Expanded Form and Classroom Edition: Report to Parents.

² VABS also has a classroom version, which is not covered here because it is for children between the ages 3 and 12.

of items and materials and the method of administration. (1) The Interview Edition, Survey Form, which is more similar in content to the original VABS, has 297 items and provides a general assessment of adaptive behavior. It is administered to a parent or caregiver in a semi-structured interview format. (2) The Interview Edition, Expanded Form, has 577 items, including 297 from the Survey Form. This form yields a more comprehensive assessment of adaptive behavior and gives a systematic basis for preparing individual educational, habilitative, or treatment programs. The Expanded Form can be used by itself, or as a follow-up to obtain more information about deficits suggested by the Survey Form.

Both versions are organized around four Behavior Domains: Communication, Daily Living Skills, Socialization, and Motor Skills.³ For the Survey Form, items are organized in domains in developmental order. For the Expanded Form, items are in clusters, which are organized in developmental order under subdomains that make up the domains.

Uses of Information: The VABS are useful in assessing an individual's daily functioning. They can be used as an evaluation and diagnostic tool

for mentally retarded individuals or individuals with other handicaps. They can also be used to develop individual educational, habilitative, and treatment programs and can monitor progress during such a program. Finally, the VABS can be used in research in which the development and functioning of handicapped and non-handicapped individuals are investigated.

Reliability: (1) Split half-reliability⁴: internal reliability tests of both the Survey and Expanded Forms were performed on caregivers of children under the age 19. The Survey Form split half coefficients for the age groups under 3 ranged from .82 to .95 for the Domains and .96 to .98 for Adaptive Behavior Composite; the Expanded Form split half coefficients ranged from .90 to .97 for the Domains and .98 to .99 for the Composite.⁵ (2) Test-retest reliability (mean of 17 days between tests): the Survey Form reliability coefficients for caregivers of children between the ages of 6 months and 2 years, 11 months ranged from .78 to .92 for the Domains and; .90 for the Adaptive Behavior Composite.⁶ There were no test-retest reliability tests for the Expanded Form. (3) Interrater reliability: the Survey Form interrater reliability coefficients, with a mean of 8 days

³VABS also has an optional maladaptive behavior domain for persons age 5 and older.

⁴The Spearman-Brown formula was used to calculate internal reliability for the domains and Guilford's formula was used for the adaptive behavior composite.

⁵The Survey Form medians for all the age groups ranged from .83 to .90 for the four domains and the median was .94 for the Adaptive Behavior Composite. The medians for the Expanded Form ranged from .91 to .95 for the domains and the median for the composite was .97.

⁶The test-retest coefficients for all age groups ranged from .81 to .86 for the domains and .88 for the composite.

between the interviews of caregivers of children ages 6 months to 18 years, 11 months, ranged from .62 to .78 for the Domains and was; .74 for the Adaptive Behavior Composite. There were no interrater reliability tests for the Expanded Form.

Validity: (1) Content validity: literature review and field tests with caregivers; (2) Criterion-related validity: The correlations between the Adaptive Behavior Composite and the original VABS unadjusted Social Quotient and Silverstein's Deviation Social Quotient, which corrects for inconsistencies in the Social Quotient, among caregivers of children between ages 6 months and 18 years, were both .55. Comparisons between the total of the raw scores for the four domains of the revised VABS and the original VABS yielded a correlation of .97 in a sample of mentally retarded adults and an age-adjusted partial correlation of .88 in a sample of hearing-impaired children. The correlation between the VABS and the Adaptive Behavior Inventory for Children, aged 5 to 11, was .58 and correlations between the revised VAB four domains and the AAMD Adaptive Behavior Scale, Part I, domains fell between .40 and .70. Correlations between VABS and the Kaufman Assessment Battery for Children (K-ABC) and the Peabody Picture Vocabulary Test-Revised (PPVT-R), two intelligence tests, ranged from .07 to .52 and .12 to .37, respectively. The differential magnitudes of these correlations is said to support the assumption that adaptive behavior scales and intelligence and achievement scales measure different areas of functioning.

Method of Scoring: Each item is rated 2 (behavior is usually or habitually performed), 1 (sometimes or partly performed), or 0 (never performed). In addition, there is a code "N" for instances when the child has never had the opportunity to perform the activity and a code "DK" when the caregiver does not know if the child performed the activity. The manuals provide users with instructions for scoring caregivers' responses. Domain and, in the Expanded Form, subdomain raw scores are obtained by summing the numerical values of the responses. Using tables in the manuals, the raw scores can be converted into standard scores (with a mean of 100 and standard deviation of 15), percentile ranks, stanines, and age equivalents. The sum of the domain standard scores is used to obtain the composite standard score. A table is then used to obtain the stanines and percentile rankings for the composite from the standard scores. The age equivalents for the composite score can be either the mean or median of the domain age equivalents. The manuals provide instructions for calculating the mean and median age equivalents. The domain standard scores are reported by age increments of 1 month up to 1 year, 11 months, and 2-month increments between 2 and 3 years. Children under 6 years old share the same standard composite scores. Computer scoring software can be purchased.

Interpretability: Each of the manuals has a chapter that provides guidelines in interpreting the assessment results and case examples. In addition, the manuals provide instructions and tables are provided for determining the statistical

significance and unusualness of: (1) differences between domain standard scores and their mean; (2) differences between pairs of domain standard scores; and (3) highest and lowest domain standard scores. The manuals also provide a five-level qualitative categorical classification system to describe children's adaptive abilities based on their domain and composite standard scores.

Training Support: The web site, www.VinelandForum.com, contains research, information, frequently asked questions, and discussion about the VABS. A training video is available to help train psychologists and others who administer the VABS Survey and Expanded Survey Form. In addition, there is a training tool available on CD-ROM designed to help professionals learn how to properly administer and score the Interview Editions of the VABS. An audiocassette,

which accompanies the Survey Form or the Expanded Form, is also available for training purposes and contains sample interviews with parents and other caregivers.

Adaptations/Special Instructions for Individuals with Disabilities: The scales are applicable to both handicapped and non-handicapped individuals.

Report Preparation Support: Parent report forms are provided for the Survey and Expanded Forms, which summarize the individual's derived scores⁷.

References:

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. *Vineland Adaptive Behavior Scales Interview Edition Expanded Form Manual*. Circle Pines, Minnesota, 1984.

⁷English and Spanish versions of the reports to parents are available.

VINELAND SOCIAL-EMOTIONAL EARLY CHILDHOOD SCALES/VINELAND SEEC, 1998

Authors:

Sara S. Sparrow, David A. Balla, Domenic V. Cicchetti

Publisher:

AGS (American Guidance Service, Inc.)

(800) 328-2560

www.agsnet.com

Initial Material Cost:

Vineland SEEC Kit: \$55 (includes Manual and 25 Record Forms)

Vineland SEEC ASSIST CD Kit: \$200

Representativeness of Norming Sample:

Standardization norms based on the normative data used to develop the Vineland Adaptive Behavior Scales. The sample included 1,200 children from birth to 5 years, 11 months selected to closely match the national norms in the 1980 U.S. census (including age, gender, geographic region, parent education, race/ethnic group, and community size).

Languages:

English (reports to parents also available in Spanish)

Type of Assessment:

Interviewer-assisted parent report

Age Range and Administration Interval: 0 through 5 years, 11 months

Personnel, Training, Administration, and Scoring Requirements:

Manual suggests that interviewers should have graduate education in early childhood development and training in interview techniques and experience in administering, scoring, and interpreting the Vineland SEEC Scales in practice sessions. Requires approximately 25 minutes to administer the assessment, and 10 to 15 minutes for scoring.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Internal Consistency and Test-Retest: 3 (.65 or higher)

Validity: 3 (.5 or higher for consistency)

Norming Sample Characteristics: 2 (older than 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Vineland SEEC Scales assess the social-emotional functioning of children from birth through 5 years, 11 months. Three scales, which combine into a Social-Emotional Composite, are used to evaluate a child's ability to pay attention, understand emotional expression, cooperate with others, construct and observe relationships, and develop self-

regulation behaviors. The three scales are: Interpersonal Relationships (44 items), Play and Leisure Time (44 items), and Coping Skills (34 Items). This assessment is administered as a semi-structured interview with the child's parent or caregiver, in which the interviewer asks general open-ended questions relating to the child's activities and behavior (these questions are designed

by the interviewer) to ascertain key developmental milestones. Since this requires that the interviewer design his or her own open-ended questions, it is critical that the interviewer has a thorough understanding of the test items and experience in conducting this type of interview.

Uses of Information: The Vineland SEEC can be used in educational and clinical settings to help identify developmental delays, plan and select appropriate activities for young children, and monitor developmental progress. The SEEC can also be used in research projects to determine the effects of various treatments or clinical interventions on young children's social-emotional functioning. The manual suggests that in order to obtain a more comprehensive description of children's development, it is useful to use the Vineland SEEC measures along with other measures of children's physical, cognitive, language and adaptive skills.

Reliability: The reliability statistics for children between 6 and 36 months indicate (1) Internal consistency (Spearman-Brown correlations): Interpersonal Relationships: .82 to .92; Play and Leisure Time: .72 to .96; Coping Skills: .87; and Composite: .89 to .97 (2) Test-retest reliability (interval ranged from 2 to 4 weeks and averaged 17 days): Interpersonal Relationships: .73; Play and Leisure Time: .74; Coping Skills: .54; and Composite: .77. (3) Inter-rater reliability (intervals ranged from 1 to 14 days and averaged 8 days): .47 to .60.

Validity: No validity studies are included in the manual for the Vineland SEEC. The manual

states that since the item content and scale structure of the Vineland SEEC replicate the socialization domain of the Vineland Adaptive Behavior Scale (ABS), its results are generalizable to the SEEC. Concurrent validity studies on the Vineland ABS indicate correlations with similar established tools range from .51 to .65.

Method of Scoring: Items in the SEEC are scored based on how often the child performs the specific activity. Options are: "usually performs," "sometimes or partially performs," "never performs," "no opportunity for the child to perform," and "don't know if the child performs." The Appendix of the SEEC contains scoring criteria for each of the scales. In addition, there is a software program--the Vineland SEEC ASSIST (Automated System for Scoring and Interpreting Standardized Tests)--available to help score and interpret the Vineland SEEC. The user can input raw scores or item scores to obtain a derived score and an interpretive report.

Interpretability: The manual contains a chapter on interpretation of scores as well as case studies on how to interpret a child's performance on the Vineland SEEC Scale. The manual includes age-based standard scores for 1-month intervals from birth through age 2, and at 2-month intervals from age 2 to age 5 years, 11 months. It is suggested that the Vineland SEEC Scale be implemented by professionals with graduate degrees and specialized training and experience in administering and interpreting early childhood assessments.

Training Support: Dr. Sara S. Sparrow (one

of the authors), is available to conduct professional workshops on effectively administering and interpreting the Vineland, but these must be specifically arranged with the publisher. Further information for users of Vineland products is available on-line at: www.vinelandforum.com.

Adaptations/Special Instructions for Individuals with Disabilities: Individuals classified in special education categories were included in the standardization sample for the Vineland SEEC Scales. The manual also directs the interviewer to derive an adjusted age for those children who were born prematurely. There is limited information in the manual regarding adaptation for children with disabilities--the manual suggests that the interviewer should follow scoring guidelines and criteria while taking into account the specific disability (for example, a child may use sign language to perform some activity described in the items on the scale).

Report Preparation Support: The Program Planning Report is meant to be used as an outline for recommended education and/or treatment plans. The Vineland SEEC ASSIST software produces reports on: personal information summary, score profile, score narrative, program planning profile, letter to parents (in English or Spanish), and recommended activities.

References:

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. *Vineland Social-Emotional Early Childhood Scales/Vineland SEEC*. Circle Pines, MN: American Guidance Service, Inc., 1998.

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. *Vineland SEEC ASSIST (Automated System for Scoring and Interpreting Standardized Tests)*. Circle Pines, MN: American Guidance Service, Inc., 2001.

WOODCOCK-JOHNSON III (WJ-III), 2001

Authors:

Richard W. Woodcock, Kevin S. McGrew, and Nancy Mather

Publisher:

Riverside Publishing

(800) 323-9540

www.woodcock-johnson.com

Initial Material Cost:

WJ III Complete Battery (Includes Cognitive Standard and Extended Test Books, Examiner's Manual, Examiner's Training Workbook, Audio Cassette, 25 Test Records and 25 Response Booklets, 5 BIA Test Records, Achievement Form A Standard and Extended Test Books, Examiner's Manual, Examiner's Training Workbook, Audio Cassette, 25 Test Records and 25 Response Booklets, WJ III Compuscore and Profiles Program (Windows and Macintosh), Technical Manual, and Scoring Guides): \$967

Achievement Battery (includes Achievement Form A or B Standard and Extended Test Books, Examiner's Manual, Audio Cassette, 25 Test Records and 25 Response Booklets, Scoring Software (Windows and Macintosh), Technical Manual, and Scoring Guides.): \$444

Cognitive Abilities Battery (includes Cognitive Standard and Extended Test Books, Examiner's Manual, Audio Cassette, 25 Test Records and 25 Response Booklets, 5 BIA Test Records, Scoring Software (Windows and Macintosh), Technical Manual, and Scoring Guides): 601

Representativeness of Norming Sample:

Normative data for the WJ-III were gathered from 8,818 people in over 100 U.S. communities. The preschool sample, (ranging in age from 2 to 5 years but not enrolled in kindergarten) included 1,143 children. The sample was selected to be representative of the U.S. population from 2 to 90 years and older.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

For 7 WJ-III Cognitive and 12 WJ-III Achievement tests, 2 to adult; for the remaining tests, school-age to adult. Achievement tests come in two equivalent forms to allow repeat administrations.

Personnel, Training, Administration, and Scoring Requirements:

Each test takes approximately 5 to 10 minutes. Examiners must provide evidence that they meet the highest standards required for using educational and psychological tests. Tests must be administered and interpreted by a highly trained individual. A computer-scoring program generates the scores.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher)

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a professional).

Description: The WJ-III consists of two assessment instruments, the Woodcock-Johnson III Tests of Cognitive Ability (WJ-III COG) and the Woodcock-Johnson III Tests of Achievement (WJ-III ACH). These instruments provide a comprehensive set of norm-referenced tests for measuring intellectual abilities and academic achievement. The instruments include a standard battery and an extended battery. The tests in the extended battery are included to allow for in-depth diagnostic testing of different aspects of an ability or achievement area. The WJ-III COG consists of a standard battery of 10 tests and an extended battery of 10 tests. Seven of the 20 COG tests are recommended for use with preschool children and can be used with children as young as 2 years of age. Those COG tests include: Test 1, Verbal Comprehension; Test 2, Visual-Auditory Learning; Test 3, Spatial Relations; Test 4, Sound Blending; Test 5, Concept Formation; Test 6, Visual Matching; and Test 17, Memory for Words. The WJ-III ACH consists of a standard battery of 12 tests and an extended battery of 10 tests. Twelve of the 22 ACH tests are recommended for use with preschool children and can be used with children as young as 2 years of age. Those ACH tests include: Test 1, Letter-Word Identification; Test 3, Story Recall; Test 4, Understanding Directions; Test 7, Spelling; Test 9, Passage Comprehension; Test 10, Applied Problems; Test 12, Story Recall-Delayed; Test 13, Word Attack; Test 14, Picture Vocabulary; Test 15, Oral Comprehension; Test 19, Academic Knowledge; and Test 21, Sound Awareness. The examiner can

tailor the administration by selecting the tests that best tap the abilities and skills of interest for a particular child. Tests can be grouped to yield meaningful summary scores.

Uses of Information: The WJ-III is designed to provide age-based or grade-based norm-referenced individual test and broad ability and achievement scores. This information can be used for a variety of purposes, including diagnosis of academic strengths and weaknesses, educational programming, assessing growth, program evaluation, and research.

Reliability: (1) Internal consistency reliability: Split-half reliabilities were calculated for all but the timed tests and tests with multiple-point scoring systems. Reliabilities for children 2 and 3 years of age ranged from .56 (for children 2 years old for Story Recall) to .98, with almost all of the correlations at the .80 level or above. The publishers recommend using cluster scores (groupings of 2 or more tests) because the cluster scores consistently have higher reliability. (2) Test-retest reliability: Studies of test-retest reliabilities for children 2 and 3 years old for the timed tests were not described in the technical manual, but for children 7 to 11 years old ranged from .76 to .94. One study of test-retest reliability over periods of 1 year or more was reported for children 2 to 7 years at first testing. For the tests appropriate for children 2 and 3 years old, reliabilities ranged from .57 (for Memory for Words 1 to 2 years after the first testing) to .96 (for Letter-Word Identification less than 1 year after the first testing)¹.

¹Note that some researchers would interpret what the authors have described as extended test-retest interval reliability as evidence of predictive validity.

Validity: (1) Content validity: The tests and clusters are based on the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. The WJ-III's content validity rests on its adherence to the CHC theory. Content was also designed to test core curricular areas and areas specified in federal legislation. Test items were developed with contributions from experts with the goal of measuring both narrow and broad abilities. Reviewers also conducted item reviews for bias to identify any items that might be potentially sensitive to women, individuals with disabilities, and cultural or linguistic minorities. Any items flagged by the reviewers were revised or deleted. The authors provided additional evidence for content validity by presenting data that demonstrate the growth and decline of cognitive and achievement abilities across the lifespan. (2) Concurrent validity: The authors also presented data demonstrating that tests from the same cluster are highly and significantly correlated and those from different clusters correlate at a lower level. For example, for children 2 to 3 years old, the Verbal Comprehension (a test of comprehension-knowledge) and Picture Vocabulary (a test of oral expression) tests are correlated at .92, but Verbal Comprehension and Concept Formation (a test of fluid reasoning) were only correlated at .32. The authors also provided evidence for the concurrent validity of the tests in a preschool sample (children ranged in age from 1 year, 9 months to 6 years, 3 months). In that study, correlations among the WJ-III standard scale generally ability tests appropriate for preschool children and the Wechsler Preschool

and Primary Scale of Intelligence-Revised ranged from .66 to .73, and with the Differential Ability Scales the correlations ranged from .57 to .67.

Method of Scoring: The general test and individual item scoring rules are summarized in the Examiner's Manuals and on the test easels (the flip books used for testing). Whether the child passes or fails an item is recorded in the test record. Raw scores are computed by summing the number of correct responses. Raw scores are entered into the computer scoring program which generates norm-referenced scores.

Interpretability: The types of scores that are available include: grade equivalents, age equivalents, relative proficiency indexes, cognitive-academic language proficiency levels, percentile ranks, and standard scores. The Examiner's Manuals provide a great deal of information about how to interpret the individual test scores, cluster scores, and discrepancies between scores in two different cognitive and ability areas.

Training Support: Training videos and workbooks are available from the publisher. The publisher offers national and regional group training sessions as well as individual training sessions. Technical support is also available by telephone and on line.

Adaptations/Special Instructions for Individuals with Disabilities: The Examiner's Manuals describe accommodations that can be made for testing young children, English language learners, and individuals with various difficulties and impairments (including reading, attention,

hearing, visual, and physical impairments).

Report Preparation Support: The Report Writer software summarizes the test findings into a clinically useful report.

References:

Mather, N. and R.W. Woodcock. Examiner's Manual: Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing, 2001.

Mather, N. and R.W. Woodcock. Examiner's Manual: Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing, 2001.

McGrew, K.S. and R.W. Woodcock. Technical Manual: Woodcock-Johnson III. Itasca, IL: Riverside Publishing, 2001.

Schrank, F.A. and R.W. Woodcock. WJ III Compuscore and Profiles Program (computer software). Itasca, IL: Riverside Publishing, 2001.

Parenting, the Home Environment, and Parent Well-Being Instruments^a



Parenting, the Home Environment, and Parent Well-Being Instruments

Instrument	Screening or Assessment	Domain			Age Range	Assessment Type	Initial Material		Ease of Administration		
		L	S-E	M			O	Cost	Reliability	Validity	Sample
Adult-Adolescent Parenting Inventory-2	S	X		X	13+ years	3	2	3	3	3	2
Beck Anxiety Inventory	S		X		17-80 years	3	1	3	3	2	2
Beck Depression Inventory-Second Edition	S		X		13-80 years	3	1	3	3	2	2
Center for Epidemiologic Studies Depression Scale	S		X		18+ years	3	1	2	3	1	2
Child Abuse Potential Inventory	S	X	X	X	Unspecified	3	2	3	2	1	2
Composite International Diagnostic Interview	A?		X		Adults	3	2	3	3	1	3
Conflict Tactics Scales, Parent-Child Version	S	X			Unspecified	3	1	2	1	1	2
Confusion, Hubbub, and Order Scale	S	X			Unspecified	3	1	3	2	1	2
Early Head Start National Evaluation Questionnaires	NA	X									
Edinburgh Postnatal Depression Scale	S		X		16+ years	3	1	1	3	1	2
Family Environment Scale	A	X			Unspecified	3	3	3	1	2	2
Family Needs Scale	S		X	X	Unspecified	3	1	3	2	1	2
Family Resource Scale	S		X	X	Unspecified	3	1	3	3	1	2
Family Support Scale	S		X	X	Unspecified	3	1	2	2	1	2
Home Observation for Measurement of the Environment	S	X	X		0-3 years	2,3	1	2	2	1	2
Infant-Toddler and Family Instrument	S	X	X	X	6-47 months	1,2	1	1	1	1	3
Kempe Family Stress Inventory	S	X	X		Unspecified	1	1	3	2	1	3
Knowledge of Infant Development Inventory	S	X	X		Unspecified	3	1	2	1	2	2
Nursing Child Assessment Teaching Scales	A	X	X		0-36 months	2	2	3	2	2	3
Parenting Stress Index	S	X	X		0-12 years	1	2	3	3	2	3
Support Functions Scale	S		X	X	Unspecified	3	1	2	2	1	2

KEY

Domains	Assessment Type	Initial Material Cost	Reliability
H = Home environment	1 = Direct assessment	1 = Under \$100	1 = None described
P = Parenting, parent-child relationship	2 = Observation	2 = \$100 to \$200	2 = Under .65
PE = Parent mental health	3 = Parent/self report	3 = More than \$200	3 = .65 or higher
F = Family functioning, support			
O = Other			

Validity

- 1 = None described
- 2 = Under .5 for concurrent; under .4 for predictive
- 3 = .5 or higher for concurrent; .4 or higher for predictive

Norming sample

- 1 = None described
- 2 = Older than 15 years, not nationally representative or representative of EHS population
- 3 = Normed within past 15 years, nationally representative or representative of EHS population

Ease of administration and scoring

- 1 = Not described
- 2 = Self-administered or administered and scored by someone with basic clerical skills
- 3 = Administered and scored by a highly trained individual

*The information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

ADULT-ADOLESCENT PARENTING INVENTORY (AAPI-2), 1999

Authors:

Stephen J. Bavolek and Richard G. Keene

Publisher:

Family Development Resources, Inc.

800-688-5822

www.nurturingparenting.com/ -or-

www.familydev.com

Initial Material Cost:

The AAPI-2 complete kit (includes the handbook, test forms A and B, Scoring Stencil for forms A and B, profiles (pkg. of 100), worksheets (pkg. Of 100): \$122. Complete kit also available on CD-ROM for \$186.

Representativeness of Norming Sample:

A non-randomly selected sample (that the authors describe as nationally representative) of adolescents and adults (abusive and non-abusive adults, abused and non-abused adolescents, and teen parents) referred by agencies from around the country using the original AAPI participated in the standardization of the AAPI-2.

Languages:

English and Spanish

Type of Assessment:

Self-report

Age Range and Administration Interval:

Persons ages 13 and older

Personnel, Training, Administration, and Scoring Requirements:

An individual can learn to score the assessment by reading the manual. No training is necessary.

Approximately 20 minutes to administer. Written at a 5th grade reading level.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher for concurrent/discriminant)

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally representative)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Adult-Adolescent Parenting Inventory (AAPI-2) is a 40-item questionnaire used to assess the parenting attitudes and child rearing practices of adolescents and adults. The purpose of the inventory is to determine the degree to which respondents agree or disagree with parenting behaviors and attitudes

known to contribute to child abuse and neglect. Responses are given on a five-point Likert scale ranging from Strongly Agree, Agree, Uncertain, Disagree, to Strongly Disagree. Responses provide a standard for risk in five parenting constructs known to contribute to the maltreatment of children: (1) inappropriate parental expecta-

tions, (2) inability to demonstrate empathy towards children's needs, (3) strong belief in the use of corporal punishment, (4) reversing parent-child family roles, and (5) oppressing children's power and independence. The AAPI-2 comes in two alternate forms—A and B—to reduce the practice effect when repeating the inventory in a short time period.

Uses of Information: Responses to the AAPI-2 permit the identification of high-risk child rearing and parenting practices that could lead to physical or emotional abuse or neglect of children. In addition, the AAPI-2 is used to (1) provide pretest and posttest data to measure

treatment effectiveness, (2) assess the parenting and child rearing attitudes of parents and adolescents prior to parenthood, (3) design specific treatment and intervention parenting education programs, (4) design nurturing experiences for parents and adolescents whose attitudes indicate a high risk for child maltreatment, and (5) screen foster parent applicants, child care staff, and volunteers for education and training purposes.

Reliability: (1) Internal reliability: Reliability coefficients for the five parenting constructs using the Spearman-Brown formula ranged from .83 to .93 on Form A, .80 to .93 on Form B, and .87 to .96 on Forms A and B combined. The Cronbach alphas ranged from .80 to .92 on both Forms A and B and .86 to .96 on Forms A and B combined.

Validity: (1) Content validity: Statements made by parents about children formed the basis of the inventory items. Professionals in the helping fields assigned items to one of the five parent-

ing constructs and assessed items' suitability for a Likert scale. (2) Construct validity: The authors provide factor analysis results that provide evidence for five underlying factors. (3) Criterion-related validity: A comparison between a group of abusive parents and a group of non-abusive parents (1,985 total sample size) found that abusive parents had mean scores on each of the parenting constructs that were statistically significantly lower than non-abusive parents. In general, males were also found to have lower scores than females, but there was no parenting-gender interaction effect. The authors provide evidence that the AAPI-2 discriminates between abusive and non-abusive parents in samples of adults and in sample of adolescents.

Method of Scoring: Scoring is completed by placing a stencil over the test items and recording the numerical value of each response (1-5 points). The numerical values are recorded on the profile worksheet for each of the five subscales. The values are summed to obtain the subscale total raw score. The AAPI-2 worksheet and table of norms located in the Handbook convert total raw scores to standard scores for developing a risk profile on the worksheet. The respondent's attitudes in each of the five sub-scales can be compared with the parenting and child rearing attitudes of parents or adolescents. Tables convert raw scores to standard ten (sten) scores by gender (male or female), parental status (parents or non-parents), and age (adults or adolescents).

Interpretability: The standard scores are plotted on the AAPI Parenting Profile, which pro-

vides an index of risk for abusive and/or neglecting behaviors. The sten scores on the Profile sheet range from 1 to 10. Low sten scores (1 to 4) generally indicate a high risk for practicing known abusive parenting practices; mid-range scores (4 to 7) represent the parenting attitudes of the general population; and high sten scores (7 to 10) indicate the expressed parenting attitudes reflect a nurturing, non-abusive parenting philosophy. The manual provides instructions for interpreting the scale's total score and the subscale scores.

Training Support: Training workshops and training assistance is available. Call 828-681-8120 or send an email to fnc@nurturingparenting.com

Adaptations/Special Instructions for Individuals with Disabilities: The assessment can be administered orally to non-readers.

Report Preparation Support: None described beyond the profiles.

References:

Bavolek, Stephen J. and Richard G. Keene. Adult-Adolescent Parenting Inventory AAPI-2 Administration and Development Handbook. Park City, UT: Family Development Resources, Inc., 1999.

BECK ANXIETY INVENTORY (BAI), 1993

Authors:

Aaron T. Beck and Robert A. Steer

Publisher:

The Psychological Corporation

(800) 228-0752

www.psychcorpcenter.com/content/bai.htm

Initial Materials:

Complete Kit: \$66 (Includes Manual and 25 Record Forms)

Spanish Record Forms (25/package): \$35

Representativeness of Norming Sample: Not nationally representative (the three normative samples of psychiatric outpatients were drawn from consecutive routine evaluations at the Center for Cognitive Therapy in Philadelphia, Pennsylvania. The total sample size was 1,086.)

Languages:

English and Spanish

Type of Assessment:

Self-report

Age Range and Administration Interval:

17 through 80 years

Personnel, Training, Administration, and Scoring Requirements:

The BAI may be administered and scored by paraprofessionals, but it should be used and interpreted only by professionals with appropriate clinical training and experience. The BAI requires 5 to 10 minutes to complete when it is self-administered and 10 minutes when it is orally administered. Scoring takes 5 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher for concurrent)

Norming Sample Characteristics: 2 (not nationally representative, normed within past 15 years)

Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

Description: The Beck Anxiety Inventory (BAI) is a 21-item scale that measures the severity of self-reported anxiety in adults and adolescents. It consists of descriptive statements of anxiety symptoms which are rated on a 4-point scale with the following correspondence: "Not at all" (0 points); "Mildly; it did not bother me much" (1); "Moderately; it was very unpleasant, but I could stand it" (2); and "Severely; I could barely stand

it" (3).

Uses of Information: The Beck Anxiety Inventory was specifically designed to reduce the overlap between depression and anxiety scales by measuring anxiety symptoms shared minimally with those of depression.

Reliability: (1) Internal consistency (Cronbach's alpha) ranged from .92 to .94 for

adults. The alphas for the Diagnostic and Statistical Manual of Mental Disorders, Third Edition—Revised (DSM-III-R) anxiety disorder groups ranged from .85 to .93. (2) Test-retest reliability (1-week interval): .75. Reliability of the BAI for adolescents has not been directly tested.

Validity: (1) Concurrent validity: the correlation with the Hamilton Anxiety Rating Scale—Revised was .51. The correlation with the anxiety subscale of the Cognition Check List, which measures the frequency of dysfunctional cognitions related to anxiety, was also .51. The BAI is also significantly correlated with the Trait (.58) and State (.47) subscales of the State-Trait Anxiety Inventory (Form Y) and with the mean 7-day anxiety rating (.54) of the Weekly Record of Anxiety and Depression. Validity of the BAI for adolescents has not been directly tested.

Method of Scoring: The BAI total score is the sum of the ratings for the 21 symptoms. Each symptom is rated on a 4-point scale ranging from 0 to 3. The maximum score is 63 points. The Beck Computer Scoring (BCS) program also scores and interprets the BAI.

Interpretability: According to the 1993

Revisions of the BAI manual, total scores of 0 to 7 reflect “Minimal level of anxiety”; scores of 8 to 15 indicate “Mild anxiety”; scores of 16 to 25 reflect “Moderate anxiety”; and scores of 26 to 63 indicate “Severe anxiety.” Because a BAI total score yields only an estimate of the overall severity of anxiety being described by a person, the clinician interpreting the score should consider other aspects of the individual’s psychological functioning.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: If an examinee needs help to complete the inventory, instructions are given in the manual for how to administer the inventory orally.

Report Preparation Support: None described.

References:

Beck, Aaron T. and Robert A. Steer. Beck Anxiety Inventory Manual. San Antonio, TX: The Psychological Corporation Harcourt Brace & Company, 1993.

BECK DEPRESSION INVENTORY—SECOND EDITION (BDI-II), 1996

Authors:

Aaron T. Beck, Robert A. Steer, and Gregory K. Brown

Publisher:

The Psychological Corporation
(800) 228-0752
www.psychcorp.com

Initial Material Cost:

Complete kit: \$65 (includes manual and 25 record forms)
Beck InterpreTrak: \$50

Representativeness of Norming Sample:

Clinical and non-clinical sample of 500 outpatients from two urban and two rural psychiatric institutes

Languages:

English and Spanish (record forms only)

Type of Assessment:

Self-report

Age Range and Administration Interval: 13 through 80 years

Personnel, Training, Administration, and Scoring Requirements:

Requires only a few minutes to familiarize oneself with the inventory, 5 to 10 minutes to administer, and a few minutes to score.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Internal consistency and test-retest reliability: 3 (.65 or higher)

Validity: 3 (Concurrent validity mostly above .5)

Norming Sample Characteristics: 2 (conducted within past 15 years, not nationally representative)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Beck Depression Inventory (second edition) is a self-administered tool for screening and assessing the severity of depression in adolescents and adults. Twenty-one items assess the intensity of depression in diagnosed patients as well as detect possible depression in normal population. Each item is a list of four statements arranged in increasing severity about a particular symptom of depression. This version of the BDI is in compliance with DSM-IV criteria for depression, and the age range covered has been expanded to 13 to 80 years of age.

Uses of Information: The BDI-II is an assessment of the severity of depression in psychiatrically diagnosed adults and adolescent patients aged 13 and older. It was developed as an indicator of the presence and degree of symptoms correlated with depression as defined in the DSM-IV (including suicidality), not as an instrument for specifying a clinical diagnosis. It is also used extensively to monitor therapeutic progress.

Reliability: (1) Internal consistency (Cronbach's alpha) is .92 for clinical patients and .93 for non-clinical individuals. (2) Test-retest

reliability is .93 (only for a small subsample of outpatients, tested with a one week lapse).

Validity: (1) Concurrent validity: two comparisons between BDI-II and its previous version resulted in correlations of .93 and .84, the latter using the take-home form. Other tests found BDI-II to be correlated with the Beck Hopelessness Scale (.68), Scale for Suicide Ideation (.37), Beck Anxiety Inventory (.60), Hamilton Psychiatric Rating Scale for Depression-Revised (.71), and Hamilton Rating Scale for Anxiety - Revised (.47).

Method of Scoring: Most items on the BDI-II are rated on a 4-point scale ranging from 0 to 3. Several items have seven response options to discern differences in behavior or motivation. The BDI-II is scored by adding the ratings for the 21 items. The maximum total score is 63.

Interpretability: The interpretation of the scores should be done by professionals who have appropriate training and experience. Clinical interpretation of total scores uses the following guidelines: 0 to 13 (minimal depression),¹ 14 to 19 (mild depression), 20 to 28 (moderate depression), and 29 to 63 (severe depression). The Beck InterpreTrak software package offers a quick analysis of results for all of the Beck Scales (Depression, Anxiety, Hopelessness, and Suicidal Ideation) and then summarizes results in a single interpretive report with insights from Aaron T.

Beck, M.D. InterpreTrak also helps monitor progress by generating longitudinal graphs and outcome ratings for each patient.

Training Support: Minimal training is required for administering or scoring the scale.

Adaptations/Special Instructions for Individuals with Disabilities: For individuals with reading or concentration difficulties, the items may be read aloud by the examiner. The manual includes instructions for both oral and self-administration. The manual also includes brief guidelines on how to help patients with severe depression understand the range of responses to the questions.

Report Preparation Support: The InterpreTrak software (available in CD-ROM or diskette) produces a comprehensive interpretive report.

References:

Beck, Aaron T., Gregory K. Brown, and Robert A. Steer. Beck Depression Inventory-II (BDI-II). San Antonio, TX: The Psychological Corporation, 1996.

Beck, Aaron T. Beck InterpreTrak. San Antonio, TX: The Psychological Corporation, 2000.

¹Note that some research has shown that women who scored 0 or 1 tend to exhibit similar behaviors to high scoring women when observed in parent-child play. This has been attributed to denial—healthy people experience and endure at least some symptoms of depression.

CENTER FOR EPIDEMIOLOGIC STUDIES DEPRESSION SCALE (CES-D), 1985

Authors:

L.S. Radloff

Publisher:

National Institute of Mental Health

Initial Material Cost:

None

Representativeness of Norming Sample: Scale not normed.

Languages:

English

Type of Assessment:

Self-report

Age Range and Administration Interval: Intended for individuals over 18

Personnel, Training, Administration, and Scoring Requirements:

No training is required. The scale takes about 10 minutes to complete, and only a few minutes to score.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Internal consistency: 3 (.65 or higher);

Test-retest: 2 (<.65)

Validity: Concurrent: 3 (.5 or higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Center for Epidemiological Studies-Depression Scale (CES-D) is a 20-item instrument that can be self-administered or administered with minimal involvement by an interviewer. The instrument was developed by the National Institute of Mental Health to detect major or clinical depression in the general (nonpsychiatric) adult population (i.e., persons older than 18), specifically the frequency and duration of cognitive, affective, and behavioral depressive symptoms (within the past week).

Uses of Information: The CES-D is used for

initial screening of symptoms related to depression or psychological distress. However, because the CES-D does not assess the full-range of depression symptoms (for example, it does not assess suicidality) and because it assesses the occurrence of the symptoms during the past week, users are cautioned against relying on the CES-D exclusively. It has also been used extensively for research purposes to investigate levels of depression among the nonpsychiatric population.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) ranged from .84 to .90 in field

studies. (2) Test-retest reliability: Ranges from .51 to .67 in 2- to 8-week intervals and .41 to .54 in 3- to 12-month intervals.

Validity: (1) Concurrent validity: studies have examined the degree to which CES-D scores are in agreement with other measures of depression. These studies found CES-D to have correlations ranging from .50s to .80s with the Hamilton rating scale, .30s to .80s with the Raskin rating scale, .40s to .50s with the Lubin Depression Adjective Checklist, .60s and .20s, respectively, with the Bradburn Affect Balance Scale's Negative Affect and Positive Affect Scales, .50s with the Langner scale and .43 with the Cantril life satisfaction ladder. Discriminant validity tests found CES-D to be less successful in differentiating between depression and other types of emotional responses, such as anger, fear, and boredom.

Method of Scoring: Respondents indicate the frequency or duration of time (in the past week) during which they have experienced certain feelings/situations. They circle a number between 0 and 3; 0 indicates that the situation occurred "rarely or none of the time" (less than 1 day), 1 indicates "some or a little of the time" (1 to 2 days), 2 indicates "occasionally or a moderate amount of time" (3 to 4 days), and 3 indicates "most or all of the time" (5 to 7 days). After adjusting the scores for the four positive-feature items, the item scores are summed to obtain the total scale score.

Interpretability: The possible range of total

scores is from 0 to 60, with higher scores indicating greater distress. Radloff, the author of the scale, suggests that a total score of 16 be used as the cutoff to indicate "case" depression.

However, other studies have suggested that scores of 0 to 15.5 be interpreted to indicate that an individual is "not depressed", 16 to 20.5 to indicate "mild depression", 21 to 30.5 to indicate "moderate depression", and 31 or higher to indicate "severe depression". It is suggested that the scale be used only as an indicator of symptoms relating to depression, not as a means to clinically diagnose depression. Therefore, higher scores on the CES-D scale may indicate a need for further clinical tests/screenings. However, because of the CES-D's limitations, a low score does not necessarily indicate the absence of clinical depression.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Devins, Gerald M. and Carolee M. Orme. "Center for Epidemiologic Studies Depression Scale." In *Test Critiques*, edited by D.J. Keyser and R.C. Sweetland. Kansas City, MO: Test Corporation of America, 1985.

Radloff, L.S. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement*, vol. 1, 1977, pp. 385-401.

THE CHILD ABUSE POTENTIAL INVENTORY (CAP), SECOND EDITION, 1986

Authors:

Joel S. Milner

Publisher:

Psytec Corporation
(815) 758-1415

Initial Material Cost:

CAP Inventory Manual: \$30
Interpretive Manual: \$20
Package of ten tests: \$16 (also available in packages of 25, 50, and 100)
Hand-scoring templates: \$50
CAPSCORE computer scoring program: \$195 (software is currently being updated)

Representativeness of Norming Sample: Not nationally representative

Languages:

English and Spanish

Type of Assessment:

Parent or caregiver self-report

Age Range and Administration Interval:

Not applicable

Personnel, Training, Administration, and Scoring Requirements:

A trained nonprofessional under the supervision of a qualified professional can administer the Inventory. However, interpretation should be done by a professionally trained social worker, counselor, psychologist, or other professional with advanced training in assessment and test interpretation. This inventory requires a 3rd grade reading level. It takes 20 minutes to administer. Scoring can be done by computer or by hand.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher) for 77-item CAP abuse scale

Validity: 2 (.5 or higher for concurrent and <.4 for predictive)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered, scored by a highly trained individual or computer program)

Description: The CAP Inventory is a 160-item questionnaire designed to assist in screening male and female parents or primary caregivers who are suspected of physical child abuse. The Inventory (Form VI) contains a total of 10 scales. The primary clinical scale is the 77-item physical child abuse scale. This abuse scale can be divided

into six factor scales: distress, rigidity, unhappiness, problems with child and self, problems with family, and problems from others. In addition, the CAP Inventory contains three validity scales: the lie scale, the random response scale, and the inconsistency scale. The validity scales are used in various combinations to produce three

response distortion indexes: the faking-good index, faking-bad index, and random response index. This instrument should always be used in conjunction with evaluation data from other sources, including interviews and other test data.

Uses of Information: The CAP Inventory is intended to assist in the screening of suspected physical child abuse cases in social services agencies and similar settings. It can also be used as a screening tool for the selection of individuals who are at increased risk for physical child abuse, to assess clients prior to treatment, or for treatment or program evaluation purposes.

Reliability: (1) Split-half reliability: Split-half reliabilities ranging from .93 to .98 and Kuder Richardson-20 coefficients ranging from .85 to .96 have been reported for different gender, age, educational level, and ethnic groups.

Validity: (1) Construct validity: The CAP abuse score is positively correlated (.48) with the amount of physical abuse in childhood. (2) Predictive validity: a significant correlation of .34 was found between abuse scores and subsequent confirmed reports of abuse and neglect.

Method of Scoring: Each item is answered in a forced-choice, agree-disagree format. Scoring can be done by hand or by computer using a computer-scoring program (CAPSCORE). The hand scoring approach uses a series of transparent scoring templates to generate the scale scores. The name of each CAP Inventory scale and the associated items to be scored are indicated on each template. For the Abuse Scale and six factor scales, weighted scores are then summed and scale scores are produced. For the Validity Scales, a

nonweighted scoring procedure is used.

The Response Distortion Index Scores are determined using the raw score totals of different pairs of the individual validity scales. A non-weighted scoring procedure is used for Special Scale Scores. To avoid errors, it is recommended that the CAPSCORE program be purchased and used to score the Inventory. It computes all of the above scores automatically.

Interpretability: Interpretation rules for the validity scales, the response distortion indexes, the CAP abuse scale, and the six abuse factor scales are outlined in the Technical and the Interpretive Manuals. Cut-off scores are listed in the manual. The author recommends using a cut-off score of 116, or 215 out of the possible 0 to 400 points, depending on the sensitivity and specificity required. The manual specifies that the 77-item abuse scale score, not individual factor scores, should be employed for the screening of physical child abusers.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Milner, Joel S. *The Child Abuse Potential Inventory: Manual (Second Edition)*. DeKalb, IL: Psytec, 1986.

Milner, Joel S. *An Interpretive Manual for The Child Abuse Potential Inventory*. Webster, NC: Psytec, 1990.

COMPOSITE INTERNATIONAL DIAGNOSTIC INTERVIEW (CIDI), 1997

Authors:

World Health Organization

Publisher:

World Health Organization www.who.int/msa/cidi/

Initial Material Cost:

Complete Package (includes lifetime and 12-month interviews, plus interviewer's and trainer's manual, Probe Flow Chart, and question specifications. Also includes the scoring and data entry programs for both interviews.): \$150; CIDI Auto (The computerized version of the CIDI): \$500

Representativeness of Norming Sample: None described

Languages:

English, Spanish, French, Italian, Dutch, German, and Portuguese ¹

Type of Assessment:

Self report

Age Range and Administration Interval:

Adults who can read and write

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 75 minutes for the regular form; 20 minutes for the short form. The instruments can be administered by a trained nonclinician and scored by a clerical individual. Computer scoring takes 20 minutes.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: Inter-rater: 3 (kappa .65 and higher), Test-retest: 3 (kappa mostly .65 and higher)

Validity: Concurrent: 3 (mostly .5 and higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Composite International Diagnostic Interview (CIDI) is a comprehensive, fully-structured psychiatric diagnostic interview designed to be used by trained nonclinician interviewers to diagnose more than 40 mental disorders among adults from different cultures according to the definitions and criteria of both the International Classification of Diseases, 10th edition (ICD-10) and the Diagnostic and Statistical

Manual of Mental Disorders, 4th edition (DSM-IV) diagnostic systems for lifetime, last year, last 6 months, last month, and last 2 weeks. The CIDI is available in lifetime and 12-month versions, and in both paper and pencil and computer-administered forms. The latter version is suitable for self-administration by cooperative subjects. During a CIDI interview, respondents are asked closed-ended questions about symptoms of psy-

¹Other language versions are in preparation.

chiatric disorders. Positive responses to some of the symptom questions are followed by questions from the Probe Flow Chart that determine whether the symptom is a possible psychiatric symptom (that is, it is clinically significant and is not due to medication, drugs or alcohol or to a physical illness or injury). Negative responses to symptom questions will often lead to later questions being skipped. If enough symptoms have been endorsed, and these symptoms occur in a pattern that suggests a diagnosis might be present, respondents are asked about the onset and the recency of the particular cluster of symptoms that they have endorsed. In addition to the CIDI, a shortened form exists, CIDI-SF, used for the period of the past 12 months. Six DSM-IV mental disorders and two DSM-III-R substance disorders are evaluated with the CIDI-SF: major depression, generalized anxiety, specific phobia, social phobia, agoraphobia, panic attack, alcohol dependence, and drug dependence.

Uses of Information: The CIDI is used to determine whether or not a person is likely to suffer from a mental disorder.

Reliability : (1) Inter-rater reliability: An intraclass kappa of 1.00. (2) Test-retest: Test-retest (with a one-month interval) kappa coefficients for substance abuse disorders over the respondents' lifetime using the Munich CIDI ranged from .55 (drug abuse) to .83 (alcohol abuse). The Brazilian CIDI yielded test-retest (no test interval provided) kappa coefficients ranging from .61 to 1.00 on all

psychiatric and substance abuse/dependency disorders, except for alcohol abuse, which had a coefficient of .35. The kappas for simple phobia, social phobia, and agoraphobia over the respondents' lifetimes were .46, .47, and .63, respectively, and, for generalized anxiety disorder, it was .53.

Validity²: Concurrent validity (referred to as "concordance validity" by authors): A comparison between the CIDI and the Structured Clinical Interview for DMS-III-R (SCID) on simple phobia, social phobia, and agoraphobia disorders yielded kappa coefficients of .45, .62, and .63, respectively, and .35 on the lifetime generalized anxiety disorder. A comparison between the CIDI and the clinical DSM-III-R criteria checklist produced kappa coefficients of .84, .83, and .76 for depressive, psychoactive substance, and anxiety disorders, respectively, and .78 for all disorders. The canonical correlation coefficients between the CIDI and the Schedules for Clinical Assessment in Neuropsychiatry on anxiety and depressive disorders were .66 for lifetime disorders and .69 for current disorders. Two studies, one that compared the CIDI-Auto with those of psychiatrists and the other that compared CIDI-Auto with pairs of clinicians, found that CIDI-Auto tended to identify more disorders than the mental professionals. The CIDI-Auto generated an average of 2.3 diagnoses of general disorders per subject compared to 1.3 diagnoses for psychiatrist and twice as many anxiety diagnoses than the clinicians. A kappa coefficient of .23 was obtained

² Many of these studies were conducted using an earlier version of the CIDI.



between the diagnoses of the CIDI and the psychiatrists. The CIDI-Auto sensitivity was above .85 for all anxiety disorders, except generalized anxiety disorder, which had a sensitivity of .29.³ Its specificity ranged from .47 to .99.⁴ The level of agreement between the CIDI-Auto and the clinicians, as measured by intraclass kappa, ranged from .02 to .81, with an overall kappa of .40.

Method of Scoring: Scoring can be done manually or by using the computerized version of the CIDI, known as CIDI-Auto, which is an SPSS-based program. The CIDI-SF is scored manually by summing the number of positive responses to symptoms the respondent reported. The manual for scoring the CIDI-SF is available at the CIDI web site.

Interpretability: No instructions were available for interpreting the CIDI. The CIDI-SF uses a probability-of-caseness score to indicate the likelihood that the respondent would meet the full diagnostic criteria if given the complete CIDI. Tables are used to convert CIDI-SF raw scores for each disorder into probability-of-caseness values that range from 0.0 to 1.0. Alternatively, the examiner can elect not to use the probability values and, instead, consider all probabilities greater than .50 to indicate that the respondent would be

a CIDI case for that disorder.

Training Support:

Administration of the interview requires training on skip patterns, on the use of the Probe Flow Chart, in assembling lists of the endorsed symptoms for the onset and recency questions, and in the use of the data entry and scoring program. Training in administering this structured interview is conducted at nine WHO endorsed centers around the world. The WHO-CIDI web site URL is www.who.int/msa/cidi/. The cost is \$1000.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: None

References:

Robins, Lee N., John Wing, Hans Ulrich Wittchen, John E. Helzer, Thomas F. Babor, Jay Burke, Anne Farmer, Assen Jablenski, Roy Pickens, Darrel A. Regier, Norman Sartorius, Leland H. Towle. "The Composite International Diagnostic Interview: An Epidemiologic Instrument Suitable for Use in Conjunction With Different Diagnostic Systems and in Different Cultures." *Arch Gen Psychiatry*, Vol. 45, Dec. 1988, pp. 1069-1077.

World Health Organization. *Composite International Diagnostic Interview (CIDI), Core Version 2.1, Interviewer's Manual*. World Health Organization, January 1997.

³ Sensitivity is a measure of the instrument's ability to correctly identify persons with the disorder as having the disorder.

⁴ Specificity is a measure of the instrument's ability to identify persons who do not have the disorder as not having the disorder.

CONFLICT TACTICS SCALES, PARENT-CHILD VERSION (CTSPC), 1998

Authors:

Murray A. Straus, Sherry L. Hamby, David Finkelhor, David W. Moore, & Desmond Runyan

Publisher:

Family Research Laboratory
University of New Hampshire
(603) 862-1888
pubpages.unh.edu/~mas2/CTS_Application.htm

Initial Material Cost:

Conflict Tactics Scale (CTS) Handbook: \$25 (This 350-page manual includes all versions of the CTS, most of the currently available CTS series publications, Spanish translations of the original CTS-Forms N and R, and other papers. The article in Child Abuse And Neglect (paper CTS17 on the above website) is the only manual for the CTSPC. However, there are other relevant articles such as CTS24 and CTS28.)

Representativeness of Norming Sample: Not normed.

Languages:

English and French translations exist for the CTSPC. Other forms of the CTS may have other translations.

Type of Assessment:

Parent report (self-administration), but can also be administered in person or over the telephone (direct parent assessment).

Age Range and Administration Interval:

Parents of children

Personnel, Training, Administration, and Scoring Requirements:

An individual with roughly a 6th grade reading level can complete the scales. The only training course is a 4-hour workshop that Dr. Straus runs every year in connection with an annual conference on family violence research. Administration time is 10 to 15 minutes if the entire scale is administered and 6 to 8 minutes if the supplemental questions are omitted.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 2 (under .65)

Validity: 1 (statistics not provided)

Norming Sample Characteristics: 1 (not described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Conflict Tactics Scales, Parent-Child Version (CTSPC) is intended to measure psychological and physical maltreatment and neglect of children by parents, as well as non-violent modes of discipline. It measures the extent to which a parent has carried out specific acts of physical and psychological aggression,

regardless of whether the child was injured. Variables are measured on three scales: Non-Violent Discipline, Psychological Aggression, and Physical Assault, as well as supplemental scales that measure Weekly Discipline, Neglect, and Sexual Abuse.

Uses of Information: The CTSPC may be used as a screening tool for child maltreatment or for evaluating prevention and treatment of physical and psychological maltreatment of children.

Reliability:¹ (1) Internal reliability (Cronbach's alphas): Overall Physical Assault Scale: .55; Psychological Aggression: .60; Nonviolent Discipline: .70; Neglect Scale: .22; Severe Physical Assault Subscale: -.02. The authors attribute the low neglect and severe assault alphas to the infrequency of the events that make up the scales, thereby reducing the likelihood for high inter-item correlations. (2) Test-retest reliability is not yet available for the CTSPC. However, the test-retest reliability coefficients on the original CTS (test interval not specified) ranged from .49 to .80.

Validity:² The authors tested for construct validity by examining the direction of the relationship between subscale scores and demographic characteristics associated with child maltreatment, such as age of parent, age of child, race/ethnicity, and gender of parent. The directions of the relationships were consistent with previous findings.

Method of Scoring: Most of the scales can be scored four ways: (1) Annual prevalence, which measures whether one or more acts in the scale occurred during past year; (2) annual chronicity, which measures the number of times an act in a

scale occurred among those who used that act; (3) ever prevalence, which measures if an act ever occurred; and (4) annual frequency, which measures the number of times an act occurred. To obtain the frequency, the midpoints for the response categories chosen by the participant are summed.

Interpretability: Normative tables for the CTSPC have not yet been developed.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Straus, M.A., S. L. Hamby, D. Finkelhor, D.W. Moore, and D. Runyan. "Identification of Child Maltreatment With the Parent-Child Conflict Tactics Scales: Development and Psychometric Data for a National Sample of American Parents." *Child Abuse & Neglect*, vol. 22, no. 4. 1998.

Straus, Murray A. "Scoring and Norms for the CTS2 and CTSPC." pubpages.unh.edu/~mas2/CTS28.pdf, 1998.

Straus, Murray A. "Child-Report, Adult-Recall, and Sibling Versions of the Revised Conflict Tactics Scales." pubpages.unh.edu/~mas2/CTS24.pdf, 1999.

¹ Previous versions of the CTS were tested for reliability and validity.

² Previous versions of the CTS were tested for reliability and validity.

CONFUSION, HUBBUB, AND ORDER SCALE (CHAOS), 1995

Authors:

Adam P. Matheny, Jr., Theodore D. Wachs, Jennifer L. Ludwig, and Kay Phillips

Publisher:

Child Development Unit
Department Pediatrics
University of Louisville Health Service Center

Initial Material Cost:

None. The scale is available in the Journal of Applied Developmental Psychology article cited below.

Representativeness of Norming Sample: Instrument is not normed.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Age of children not specified, but the assessment appears targeted for homes with infants and toddlers.

Personnel, Training, Administration, and Scoring Requirements:

A caregiver or parent who is literate can answer the 15 items. Scoring requires summing the responses given by the parent and takes under 5 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher)

Validity: 2 (concurrent under .5)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The CHAOS scale is a questionnaire filled out by parents that is designed to assess the level of confusion and disorganization in the child's home environment. The questionnaire consists of 15 statements, to each of which a parent or caregiver assigns a number between 1 and 4 that correspond to the following: 1 = Very much like your own home; 2 = Somewhat like your own home; 3 = A little bit like your own

home; 4 = Not at all like your own home.

Uses of Information: The CHAOS scale screens for a chaotic home environment. High levels of chaos for at-risk children may warrant a more detailed environmental assessment to determine how and to what degree ongoing chaos is either compounding the effects of existing biosocial risks or attenuating the impact of corrective intervention.

Reliability:¹ (1) Internal consistency (Cronbach's alpha): For the entire scale, .79. (2) Test-retest reliability (12-month interval): for the total test score, .74.

Validity:² (1) Concurrent validity: the CHAOS scale was compared with the physical and social environment codes in the Purdue Home Simulation Inventory (PHSI), which are completed by trained observers. The authors report that the correlations between the CHAOS scale and several of the PHSI social environment codes were significant (physical interference (correlation = -.36), number of known objects named (correlation = -.38), and ignores bids (correlation = .45)), and together, the PHSI social environment codes explained 59 percent of the variance in the CHAOS scores. The correlations between the CHAOS scale and several of the PHSI physical environment codes also were significant (number of siblings (correlation = .55) and number of rooms per person (correlation = -.33)), and together the PHSI physical environment codes explained 39 percent of the variance in the CHAOS scores.

Method of Scoring: The statements are scored using a 4-point scoring system. A single score is derived from the CHAOS questionnaire by summing the responses for the 15 items. A higher score represents characteristics of a more chaotic, disorganized, and hurried home.

Interpretability: The higher the score, the more chaotic a home is considered to be.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Matheny, Adam P., Jr., Theodore D. Wachs, Jennifer L. Ludwig, and Kay Phillips. "Bringing Order Out of Chaos: Psychometric Characteristics of the Confusion, Hubbub, and Order Scale." *Journal of Applied Developmental Psychology*, vol. 16, 1995, pp. 429-444.

¹ These results are based on an earlier version of the CHAOS Scale that used a true-false scoring system.

² These results are based on an earlier version of the CHAOS Scale that used a true-false scoring system.

EDINBURGH POSTNATAL DEPRESSION SCALE, 1987

Authors:

J.L. Cox, J.M. Holden, and R. Sagovsky

Publisher:

None.

Initial Material Cost:

A copy of the scale can be found at www.clinical-supervision.com/edinburgh%20scale.htm

Representativeness of Norming/Research Sample:

No norming sample

Languages:

English

Type of Assessment:

Self-report

Age Range and Administration Interval:

Women of childbearing age

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 10 minutes; Scoring can be done in about 5 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 3 (.5 or higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Edinburgh Postnatal Depression Scale (EPDS) is a measurement tool that is used to screen for depression during the postpartum (postnatal) period. Mothers underline the response items that most closely reflect her feelings during the past week.

Uses of Information: The EPDS is designed to detect women suffering from postnatal depression. It does not provide information on the severity of the depression. A respondent whose score is indicative of probable postnatal depression should have a comprehensive assessment.

Reliability: None described.

Validity: (1) Concurrent Validity: a valida-

tion study on British mothers found that a 12.5 cutoff score identified over 80 percent of the mothers with major depression and about 50 percent of the mothers with minor depression, and had a sensitivity value of 67.7 percent. Another study found a score of 9.5 or higher to be more appropriate for identifying depression among Chinese mothers.

Method of Scoring: Responses are scored from 0 to 3 according to increased severity of the symptoms. Individual items are totaled to give an overall score.

Interpretability: A score of 12 or more on EPDS or an affirmative answer on question 10

(presence of suicidal thoughts) requires more thorough evaluation.

Training Support: None described, but none seems to be needed.

Adaptations/Special Instructions for Individuals with Disabilities: None described. However, the instrument can quite easily be administered in an interview format, if mental or physical disabilities make it difficult for a respondent to complete the instrument.

Report Preparation Support: None described.

References:

Lee, DTS, and TKH Chung. "What should be done about postnatal depression in Hong Kong?" *Hong Kong Medical Journal*, Vol. 5, No. 1, pp. 39-42, March 1999.

Murray, Lynne and Andrew D. Carothers. "The Validation of the Edinburgh Post-natal Depression Scale on a Community Sample." *British Journal of Psychiatry* Vol. 157, pp. 288-290, 1990.

Warner, R., Appleby, L., Whitton, A., & Faraghen, B. "Demographic and obstetric risk factors for postnatal psychiatric morbidity." *British Journal of Psychiatry*, Vol. 168, 607-611, 1996.

FAMILY ENVIRONMENT SCALE (FES), 1994

Authors:

Rudolf H. Moos and Bernice S. Moos

Publisher:

Mind Garden

(650) 261-3500

www.mindgarden.com;

Initial Material Cost:

Manual: \$56

Interpretative Report Forms: \$1 each or \$34 for 25

FES Item Booklets, Expectations and Ideal Forms: \$2 each or \$48 for 25; Real Form: \$1 each or \$32 for 25

Scoring Key: \$15 each

Self-Scorable Answer Sheets: \$1 each or \$41 for 25

Non-Paid Answer Sheets: \$1 each or \$16 for 25

Self-Scorable Preview Kit: \$57

Representativeness of Norming Sample: Form R was normed on a sample of 1,432 normal families and 788 distressed families. The normal families were diverse in terms of geography, family type, race, and age. They also included 601 families that served as normal comparison groups in studies of alcoholic and depressed families.¹ The distressed families had members who were alcohol abusers, depressed or psychiatric patients, family clinic patients, on probation or parole, and adolescents or younger children in crisis situations. Form I was normed on a sample of 591 individuals from varied family types, including normal and distressed individuals. No separate norming sample was drawn for Form E. Form E scores are normed using the Form R sample. Cross-cultural normative samples are available from the translated and culturally adapted versions of the FES.

Languages:

English, Arabic, Chinese, Dutch, Estonian, French, German, Hebrew, Hindi, Italian, Japanese, Korean, Marathi, Portuguese, Russian, Spanish, and Swedish.

Type of Assessment:

Child (11 years or older) and parent report on family environment.

Age Range and Administration Interval: Not applicable. Focus is on the family environment.

Personnel, Training, Administration, and Scoring Requirements:

Instructions for administering the FES are self-explanatory and no training is required. It usually takes individuals 15 to 20 minutes to complete each of the three forms. It takes approximately 10 minutes to obtain raw scores for all 90 items.

Summary

Initial Material Cost: 3 (>\$200)

Reliability: 3 (.65 or higher).

Validity: 1 (concurrent not available).

Norming Sample Characteristics: 2 (normed within past 15 years; diverse but not representative).

Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

¹ See description section below for information on the different types of forms.

Description: The Family Environment Scale (FES) is one of 10 Social Climate Scales, each assessing the climate in a different setting with 10 subscales organized into three dimensions—relationship, personal growth, and system maintenance. The FES measures family social environment using three forms with 90 true-false items: (1) the Expectations Form (Form E) for information on expectations from a new family environment, (2) the Real Form (Form R) for information on perceptions of the current family environment, and (3) the Ideal form (Form I) for information on the preferred family environment. In addition, there is a 30-item pictorial children's version for use with children between the ages of 5 and 11. The FES is administered to family members as a paper- and pencil-inventory with true or false answers.

Uses of Information: The FES can be used for the following purposes: (1) understanding problems in family functioning, (2) serving as a benchmark to evaluate the impact of an intervention, (3) providing feedback to families as a means to promote change, (4) evaluating how a family has been affected by a transition, life crisis or change (provided comparable information is available about the family prior to the event), (5) appraising and improving the family climate parents create, (6) strengthening families as cohesive units, (7) identifying risks for various problems, such as, depression, substance abuse, or family violence.

Reliability: (1) Internal consistency (Cronbach's alpha): the subscale alphas for Form

R ranged from .61 for independence to .78 for cohesion, intellectual-cultural orientation, and moral-religious emphasis. No alphas were reported for Form I; however, the authors reported that they were similar to Form R alphas. (2) Test-retest reliability: The Form R subscale reliability coefficients ranged from .68 for independence to .86 for cohesion with a 2-month testing interval and .54 for independence to .91 for moral-religious with a 4-month testing interval.

Validity: The authors reported, as evidence of construct validity, studies that found results on the FES subscales to be consistent with the results on other instruments measuring the same construct and the lack of a relationship with results on instruments measuring different constructs. These instruments included the Social Support Appraisals (SS-A; Vaux et al., 1986), the Social Support Questionnaire (Sarason, et al., 1987), the Locke-Wallace Marital Adjustment Scale (Waring et al., 1981), the Spanier Dyadic Adjustment Scale (DAS; Abbott & Brody, 1985), the Parental Bonding Instrument (Sarason, et al., 1987), the Family Assessment Device (FAD) and the Family Adaptability and Cohesion Evaluation Scales (FACES-II; Dickerson and Coyne, 1987), FACES-III (Edman, Cole, and Howard, 1990), the Structural Family Interaction Scale – revised (Perosa and Perosa, 1990), and the Family System Test (FAST; Feldman and Gehring, 1988), the Family Sculpture Test, and an adapted version of the Bowerman and Bahr Identification Scale (Russell, 1980). The authors did not report any statistics on the magnitude of the relationships.

Method of Scoring: Individuals complete subscale questions with true or false answers on separate answer sheets by placing an X in the appropriate column. The examiner then uses a template to score the responses by summing the number of X's in each column. When the FES is administered to more than one family member, the subscale raw scores for each family member are averaged in order to obtain the family's mean raw score for each subscale. Tables are provided to assist the clinician in converting the raw score to a standard score.

Interpretability: Subscale responses are compared to those of a group of normal families, using standard scores, which have a mean of 50 and standard deviation of 10. The manual provides tables for converting Form R subscale and family incongruence raw scores into standard scores. The same table can be used to convert Form E raw scores into standard scores. Program staff with a basic knowledge of statistics can calculate equivalent percentiles based on the family's standard scores, the mean, and the standard deviation. The manual provides case studies to help interpret the results.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: It is recommended that the FES be administered using tape-recorded or computerized instructions for those individuals who have short attention spans or cannot read at a sixth-grade level. It may also be helpful to administer the FES in individual interviews for poor functioning residents of treatment or residential care facilities. Some people will not be able to understand the questions, including chil-

dren under the age of 11 and individuals who are mentally retarded, seriously impaired psychiatrically, or who suffer from a chronic brain disorder or cognitive dysfunction.

Report Preparation Support: A sample narrative report is provided in order to help clinicians interpret the results from the FES.

References:

Abbott, D.A. & G.H. Brody. The relation of child age, gender and number of children to the marital adjustment of wives. *Journal of Marriage and the Family*, vol. 47, 1985, pp. 77-84.

Dickerson, V.C. and J.C. Coyne. Family cohesion and control: A multitrait/multimethod study, *Journal of Marital and Family Therapy*, vol. 13, 1987, pp. 275-285.

Feldman, S.S. and T.M. Gehring. Changing perceptions of family cohesion and power across adolescence, *Child Development*, vol. 59, 1988, pp. 1034-1045.

Moos, Rudolf and Bernice Moos. *Family Environment Scale Manual: Development, Applications, Research*, Third Edition. Palo Alto: Consulting Psychologists Press, Inc., 1994.

Perosa, L.M. and S.L. Perosa. Convergent and discriminant validity for family self-report measures, *Educational and Psychological Measurement*, vol. 50, 1990, pp. 855-868.

Russell, C. A methodological study of family cohesion and adaptability, *Journal of Marital and Family Therapy*, vol. 6, 1980, pp. 459-470.

Sarason, B.R., et al. Interrelations of social support measures : Theoretical and practical implications. *Journal of Personality and Social*

Psychology, vol. 52, 1987, pp. 813-832.

Vaux, A., et al. The Social Support Appraisals (SS-A) Scale : Studies of reliability and validity. American Journal of Community Psychology, vol. 14, 1986, pp. 195-219.

Waring, E.M., et al. Dimensions of intimacy in marriage. Psychiatry, vol. 44, 1981, pp. 169-175.

FAMILY NEEDS SCALE (FNS)

Authors:

Carl J. Dunst, Carolyn S. Cooper, Janet C. Weeldreyer, Kathy D. Snyder, and Joyce H. Chase

Publisher:

Brookline Books
617-558-8010, 800-666-BOOK
www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling and Empowering Families, \$25
Scales (10): \$10

Representativeness of Norming Sample: No norming sample.

Languages:

English

Type of Assessment:

Parent self-report or report by other family member

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent (or other family member) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher) for internal and split-half reliability

Validity: 2 (<.5 for total score)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Family Needs Scale (FNS) is a 41-item scale that measures a family's needs in nine areas (financial, food and shelter, vocation, child care, transportation, communication, etc.). The parent (or family member) rates each item of need on a 5-point scale that ranges from (1) almost never a need to (5) almost always a need.

Uses of Information: The scale facilitates the identification of family needs and strengths. The results can be used to guide follow-up discussions to help clarify concerns and help define the precise nature of the family's needs. The informa-

tion can then be used to decide whether intervention is required and, if so, the type of intervention needed. The book provides a number of case studies to illustrate how the information can be used.

Reliability: The research sample consisted of 54 parents of pre- and elementary school aged children who were mentally challenged, handicapped, and developmentally at risk. (1) Internal reliability: coefficient alpha = .95; (2) Split-half reliability (using the Spearman-Brown formula) = .96.

Validity: The authors tested for concurrent

validity against a parent belief scale (Snyder et al. 1986). The FNS total scale score was found to be significantly related to the well-being (correlation = .42), decision-making (correlation = .40), and internal locus of control (correlation = .28) dimensions on the parent belief scale.

Method of Scoring: The parent (or family member) reports the constancy of a need by marking Not Applicable, Almost Never, Seldom, Sometimes, Often, and Almost Always for each item of need.

Interpretability: Items rated Sometimes, Often, or Almost Always (a need) may indicate needs that are generally unmet, and thus provide a basis for further discussion to better understand the exact nature of the need. The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to record providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

Training Support: None described. However, a second book published by Brookline

Books, entitled “Supporting and Strengthening Families: Methods, Strategies, and Practices”(Dunst et al. 1994) is a collection of papers updating the thinking and practices described in “Enabling and Empowering Families,” and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Dunst, Carl, Carol Trivette, and Angela Deal. *Enabling and Empowering Families: Principles & Guidelines for Practice*. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. *Supporting and Strengthening Families: Methods, Strategies, and Practices*. Newton: Brookline Books, 1994.

Snyder, K.D., J.C. Weeldreyer, C.J. Dunst, and C.S. Cooper. *Parent Self-Awareness Scale: Reliability and Validity*. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986.

FAMILY RESOURCE SCALE (FRS), 1986

Authors:

Carl J. Dunst and Hope E. Leet

Publisher:

Brookline Books

(617) 558-8010, (800) 666-BOOK

www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling and Empowering Families, \$25

\$10 per batch of 10 scales.

Representativeness of Norming Sample: No norming sample.

Languages:

English

Type of Assessment:

Parent self-report or report by other family member

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent (or other family member) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher) for internal, split-half, and test-retest reliability

Validity: 3 (.5 or higher for criterion validity for total score)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The 31-item self-report Family Resource Scale (FRS) measures the adequacy of a family's tangible and intangible resources using a five-point scale, ranging from (1) not at all adequate to (5) almost always adequate. The scale covers such resources as food, shelter, financial resources, transportation, health care, time to be with family, child care, and time for self; which are generally organized from the most to the least essential resource. A modified version of the scales for teenage mothers is available.

Uses of Information: This scale determines the extent to which different types of resources are adequate in the households of young children. The lack of resources may be barriers to the family's involvement in their child's program, as families with unmet basic needs may not have time or energy to participate actively in the child's program. The book provides a number of case studies to illustrate how the information can be used.

Reliability: The research sample consisted of 45 mothers of preschool retarded, handicapped,

and developmentally at-risk children participating in an early intervention program. (1) Internal reliability (Cronbach's alpha): .92; split-half reliability (using the Spearman-Brown formula): .95 (2) Test-retest reliability (2 to 3 month interval): .52.

Validity: (1) Concurrent validity: both the personal well-being (Dunst 1986a) and maternal commitment (Dunst 1986b) measures were significantly related to the total scale score (.57 and .63, respectively).

Method of Scoring: The parent (or family member) marks the extent to which each of the resources is adequate for his/her family by selecting one of the following responses: Does Not Apply, Not At All Adequate, Seldom Adequate, Sometimes Adequate, Usually Adequate, and Almost Always Adequate.

Interpretability: Those items rated Not At All Adequate or Seldom Adequate may be evidence that these needs are not being met. They can provide a basis for exploring with the family the absence and need for these resources.

Training Support: None. However, a second book published by Brookline Books, entitled *Supporting and Strengthening Families: Methods, Strategies, and Practices* is a collection of papers updating the ideas and practices described in *Enabling and Empowering Families*, and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to record providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

References:

Dunst, C.J. A Short Form Scale for Measuring Parental Health and Well-Being. Unpublished manuscript. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986a.

Dunst, C.J. Measuring Parent Commitment to Professionally-Prescribed, Child-Level Interventions. Unpublished manuscript. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986b.

Dunst, Carl, Carol Trivette, and Angela Deal. *Enabling and Empowering Families: Principles & Guidelines for Practice*. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. *Supporting and Strengthening Families: Methods, Strategies, and Practices*. Newton: Brookline Books, 1994.

FAMILY SUPPORT SCALE (FSS), 1986

Authors:

Carl J. Dunst, Carol M. Trivette, and Vicki Jenkins

Publisher:

Brookline Books

617-558-8010

800-666-BOOK

www.brooklinebooks.com

Initial Material Cost:

Book: Enabling and Empowering Families, \$25

\$10 per batch of 10 scales

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Parent self-report

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 2 (.65 or higher for internal and split-half reliability; <.65 for test-retest reliability)

Validity: 2 (<.5 for criterion validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The self-report Family Support Scale (FSS) measures parents' satisfaction with the support they receive in raising a young child. The scale consists of 18 items covering such sources of support as the immediate family, relatives, friends and others in the family's social network, social organizations, and specialized and generic professional services. In addition, the scale provides 2 open items for parents to assess other sources of support not included in the 18 items. The parent rates each source of support on a 5-point Likert scale (ranging from not at all

helpful (1) to extremely helpful (5)).

Uses of Information: Providers can use the FSS scale results to identify the areas in a family's support network that need to be strengthened or accessed to better meet the families' needs. The results can also be used to initiate inquiries into issues related to the support network. The FSS might be useful as a pretest/posttest measure of perceived helpfulness of the program to the family (in relation to the family's level of involvement in the program). The book provides a number of case studies to illustrate how the information can

be used.

Reliability: (1) Internal consistency reliability: coefficient alpha (on the 18-item scale) = .77; (2) Split-half reliability (using the Spearman-Brown formula): .75 (3) Test-retest reliability (1 month interval): correlation was .75 for the average correlation among the 18 scale items and .91 for the total scale scores. Test-retest reliability (18 month interval): correlation was .41 for the 18 scale items and .47 for the total scale scores.

Validity: (1) Criterion validity: The authors compared the results on the FSS scale to results on the Parent-Child Play Scale (Dunst 1986) and selected subscales on the Questionnaire on Resources and Stress (Holroyd 1985). The FSS total scale score was consistently, but weakly, related to a number of parent and family outcomes, including personal well-being (correlation = .28), the integrity of the family unit (correlation = .18), parent perceptions of child behavior (correlation = .19), and opportunities to engage in parent-child play (correlation = .40) (Dunst 1985).

Method of Scoring: The parent answers how helpful various sources of support have been in terms of raising his/her child(ren) by circling Not Available, Not At All Helpful, Sometimes Helpful, Generally Helpful, Very Helpful, and Extremely Helpful.

Interpretability: A parent's responses are used to open up discussion as to why they use or do not use various means of support and resources. The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs

and support form can be used to record the names of providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

Training Support: None described. However, a second book published by Brookline Books, entitled "Supporting & Strengthening Families: Methods Strategies and Practices" is a collection of papers updating the thinking and practices described in "Enabling and Empowering Families," and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Dunst, C.J. "Rethinking Early Intervention." Analysis and Intervention Developmental Disabilities, vol. 5, 1985, pp. 165-201.

Dunst, C.J. A Rating Scale for Assessing Parent-Child Play Opportunities. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986.

Dunst, Carl, Carol Trivette, and Angela Deal. Enabling and Empowering Families: Principles & Guidelines for Practice. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. Supporting and Strengthening Families: Methods,

Strategies, and Practices. Newton: Brookline Books, 1994.

Holroyd, J. Questionnaire on Resources and Stress Manual. Unpublished scale. Los Angeles:

University of California, Neuropsychiatric Institute, Department of Psychiatric and Behavioral Sciences, 1985.

HOME OBSERVATION FOR MEASUREMENT OF THE ENVIRONMENT (HOME) INVENTORY ADMINISTRATION MANUAL, THIRD EDITION, 2001

Authors:

Bettye M. Caldwell and Robert H. Bradley

Publisher:

University of Arkansas

www.ualr.edu/~crtldept/home4.htm

(501) 565-7627

Cost:

Administration and scoring manual: \$30

50 Infant and toddler scoring sheets: \$9

A more in-depth manual was being prepared at the time of this review.

Representativeness of Norming Sample: None described.

Languages:

English

Type of Assessment:

Observation, supplemented by direct parent interview

Age Range and Administration Interval: Infant and toddler inventory for birth to 3 years old. Other inventories are available for children ages 3 to 15.

Personnel, Training, Administration, and Scoring Requirements:

Formal training is not required, but is recommended. It takes between 45 and 90 minutes to administer the inventory.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Internal reliability: 3 (.65 and higher), internal consistency: 2 (mostly <.65)

Validity: 2 (<.5 for concurrent), 3 (mostly .4 or higher for predictive)

Norming Sample Characteristics: 1 (none described)

Administration and Scoring: 2 (administered by a highly trained individual, scored by someone with basic clerical skills)

Description: The Home Observation for Measurement of the Environment (HOME) is designed to measure the quality and extent of stimulation available to a child in the home environment. The HOME serves as a screening

device for identifying environments that are not stimulating to children. HOME has separate inventories for infants and toddlers (birth to 3 years old), early childhood (ages 3 to 6), and middle childhood (ages 6 to 10).¹ The infant and

¹ There is a fourth inventory for early adolescence (ages 10 to 15). This summary focuses on the infant and toddler inventory. Although their items and subscale topics differ, all three instruments share a similar structure and have similar administration and scoring protocols.

toddler inventory is comprised of 45 items organized into 6 subscales: (1) responsiveness to parent, (2) avoidance of restriction and punishment, (3) organization of the environment, (4) appropriate play materials, (5) parental involvement, and (6) variety in daily stimulation. The items were selected to provide information from the child's perspective on stimuli that have been found to affect children's cognitive development. The information is collected from observations, supplemented by parent interview, during home visits that are scheduled when the child is awake and engaged in activities typical for that time of the day. A Supplement to the HOME for Impoverished Families (SHIF) has been developed to better assess the quality of the home environments of young children living in poor urban homes. The 20-item SHIF takes approximately 8 minutes to score and uses the same scoring procedures as the HOME. The SHIF should be used in conjunction with the total HOME rather than as an independent assessment.

Uses of Information: The HOME inventory is intended to identify environments that do not stimulate the cognitive development of children and to assist in the development of interventions that benefit both the caregiver and the child.

Reliability: (1) Internal consistency reliability: Cronbach's alphas were .84 for the HOME inventory and ranged from .49 to .78 for the its 6

subscales²; Kuder-Richardson coefficients were .89 for the inventory and ranged from .44 to .89 for the subscales; On a non-representative sample of poor urban families, .80 and .63 Kuder-Richardson coefficients were obtained for the HOME and SHIF, respectively. (2) Test-retest reliability: Pearson correlation coefficients were .62 for the inventory and ranged from .29 to .62 for the subscales administered to children at ages 6 and 12 months; .64 for the inventory and ranged from .27 to .64 when administered at ages 6 and 24 months; and .77 for the inventory and ranged from .30 to .77 when administered at ages 12 and 24 months. The intraclass correlation, which measures stability by comparing the similarity of paired scores relative to the total variation of all scores, resulted in slightly lower values. The intraclass correlation coefficients were .57 for the inventory and ranged from .23 to .57 for the subscales when administered at ages 6 and 12 months, .58 for the inventory and ranged from .25 to .58 for the subscales at ages 6 and 24 months, and .76 for the inventory and .30 to .76 at ages 12 and 24 months. Because HOME scores rose over time, the authors considered the lower intraclass coefficients to be a more accurate reliability measure. (3) Inter-rater reliability: The kappa statistics for inter-rater reliability ranged from .76 to 1.0 for the HOME and .79 to 1.0 for the SHIF.

Validity: Concurrent and Predictive:

² The analysis was performed on the 45 scale items that were part of a longer (72 items) version that was administered to a sample of 232 families in Syracuse, New York.

Families' HOME inventory scores administered when the child was 6, 12, and 24 months old were compared to the child's scores on the Bayley Scales of Infant Development Mental Development Index (MDI) at 6 and 12 months, the Stanford-Binet at 36 and 54 months, and the Illinois Test of Psycholinguistic Abilities (ITPA) at 37 months. HOME was found to be a better predictor of intelligence than socio-economic measures and was a stronger predictor for females and whites. (1) Comparison with the Bayley MDI: The correlations between the HOME inventory score at 6 months and the Bayley MDI at 6 and 12 months were .14 and .16 (subscale correlations ranged from .01 to .27), respectively. The correlation between the HOME at 12 months and the Bayley MDI score at 12 months was .30 (subscales ranged from .01 to .28). (2) Comparison with the Stanford-Binet: The correlations between the HOME inventory score at 6 months and the Stanford-Binet at 36 and 54 months were .50 (subscales ranged from .24 to .41) and .44 (subscales ranged from .10 to .44), respectively. The correlation between the HOME at 12 months and the Stanford-Binet at 36 months was .58 (subscales ranged from .24 to .56), respectively. The correlations between the HOME at 24 months and the Stanford-Binet at 36 and 54 months were .71 (subscales ranged from .41 to .64) and .57 (subscales ranged from .28 to .56), respectively. (3) Comparison with the ITPA: The correlations

between the HOME inventory scores at 6 and 24 months and the total ITPA score at 37 months were .39 and .61, respectively. (4) Comparison with SHIF: The Pearson correlation between the HOME and the SHIF was .69. (5) Comparison with the Nursing Child Assessment Feeding Scale (NCAFS) and the Nursing Child Assessment Teaching Scale (NCATS): In a non-representative sample of impoverished urban families, the Pearson correlation coefficients were .55 and .42 between the HOME and the NCAFS and NCATS, respectively, and .49 and .36 between the SHIF and the NCAFS and the NCATS.

Method of Scoring: The home visitor/interviewer enters a plus (+) for each item if the behavior is observed or reported and a minus (-) if it is not. Subscale and total inventory scores are derived by counting the number of pluses.

Interpretability: The summary sheet provides the scores that fall in the lowest quartile, the middle half, and the upper quartile. Homes with scores in the lowest quartile are considered to be environments at increased-risk. However, the authors also recommend that the interviewer collect other information on the child and the family to provide a context for interpreting the HOME scores. In addition, the interviewer should pay attention to patterns across the subscales since these patterns may provide information that can be used in developing and structuring the intervention. No information was provided on inter-

preting SHIF scores. However, on the sample of poor urban families, the mean SHIF score was 15.5 with a standard deviation of 2.8.³

Training Support: Workshops are offered by the authors and other trainers and videotapes of a skilled visitor are available. Information on the workshop and the videotapes are available from the authors.

Adaptations/Special Instructions for Individuals with Disabilities: The manual discusses circumstances under which it may be appropriate to use the standard version of the HOME to best describe the family environment of a child with disabilities. However, there is also a special form for families of children with disabilities, which will be discussed in the manual that is currently being written.

Report Preparation Support: A summary sheet for recording background information on the family and subscale and total inventory scores is available.

References:

Caldwell, Bettye M., and Robert H. Bradley. Home Inventory Administration Manual, Third Edition, 2001. Little Rock, AR: University of Arkansas at Little Rock, 2001.

Caldwell, Bettye M., and Robert H. Bradley. Psychometric Characteristics. Unpublished manuscript.

Ertem, Ilgi Ozturk, Brian William Cameron Forsyth, Abraham Joseph Avni-Singer, Lisa Kendall Damour, and Dominic V. Cicchetti. "Development of a Supplement to the HOME Scale for Children Living in Impoverished Urban Environments." *Developmental and Behavioral Pediatrics*, vol. 18, no.5, October 1997, pp. 322-328.

³ The standard deviation provides information on the distribution of the scores. About a third (34 percent) of the scores fall within one standard deviation above the mean (18.3) and an equal proportion fall within one standard deviation below the mean (12.7).

INFANT-TODDLER AND FAMILY INSTRUMENT (ITFI), 2001

Authors:

Nancy H. Apfel and Sally Provence

Publisher:

Paul H. Brookes Publishing Co.

(800) 638-3775

www.brookespublishing.com

Initial Material Cost:

ITFI instrument and manual package: \$45

Representativeness of Norming Sample:

Not normed. Field test involved 55 Connecticut families with 59 children ages 6 to 36 months.

Languages:

English

Type of Assessment:

Direct parent and child assessment; structured child observation (parent report if observation not possible)

Age Range and Administration Interval: 6 months to 3 years

Personnel, Training, Administration, and Scoring Requirements:

Designed for home visitors. Can also be used by family practitioners, including paraprofessionals, who have varying levels of training. Preparation sessions for orientation and instructional purposes are strongly recommended.

The Caregiver Interview and Developmental Map can be administered in two 45- to 60-minute sessions.

One 45- to 60-minute session is needed to share findings and develop a follow-up plan for the family.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics:

1 (none described)

Ease of Administration and Scoring:

3 (administered and scored by family service workers)

Description: The ITFI helps family service providers assess the well-being of children 6 months to 3 years of age and their families. It consists of four sections: the Caregiver Interview, Developmental Map, Checklist for Evaluating Concern, and the Plan for the Child and Family. The Caregiver Interview has 35 items related to

home and family life, child health and safety, and family issues and concerns. The interview is conducted with the parent(s) or primary caregiver(s). The Developmental Map is an observation of infant-toddler behavior in the areas of gross and fine motor development, social and emotional development, language development, and coping

and self-help development. It involves informally observing the child's behavior and interaction with others, as well as playing with the child using an established set of testing materials. The Checklist for Evaluating Concern is completed by the interviewer after the visit in order to assess the family in the areas of home and family environment; child health, development and safety; and stressors in the child's life. The interviewer rates concerns on a scale of 1 (low) to 10 (high) and prioritizes concerns based on these ratings. The Plan for the Child and Family determines what steps may need to be taken in order to address the interviewer's concerns about the needs of the family.

Uses of Information: Family service providers can use the ITFI as a supplement to other assessments of child and family strengths and needs, specific child symptoms and stressors, and the caregiver's ability to meet their child's basic needs. It also helps service providers work with families to develop a support plan for meeting their needs.

Reliability: None described.

Validity: None described.

Method of Scoring: Scoring is done on the three-part Checklist for Evaluating Concern, after the family service provider leaves the family's home. The checklist summarizes the family provider's impressions of family and child needs and strengths based on information from the Caregiver Interview, the Developmental Map, and observations of the caregiver-child interaction and the home environment. For each item in the

checklist, the provider indicates whether the condition is present, is of concern, or if the provider is unsure of its presence.

Interpretability: The interviewer uses the Checklist for Evaluating Concern Summary Sheet to rate the level of concern for the child and family from 1 (low) to 10 (high) and to list the family's strengths and weaknesses. The ratings of concern from the Checklist for Evaluating Concern may determine how detailed a support plan is, when it is put into action, and how intense services should be. After the service provider has considered all of the information from the summary sheet, he or she is able to prioritize the concerns while keeping the child's health, development, and safety as a top priority. The manual provides general guidelines and case studies on how to use information collected with the ITFI that can be used to prioritize needs and develop service plans. The guidelines are kept general to allow agencies and programs to incorporate the ITFI into their own protocol, purpose, and service.

Training Support: "Brookes on Location" professional development seminar, Using ITFI to Evaluate Young Children and Their Families, is available through the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Case study examples are provided in which a case is presented, family strengths and vulnerabilities are highlighted, and an action plan for intervention is developed.

References:

Apfel, Nancy and Sally Provence. Manual for the Infant-Toddler and Family Instrument (ITFI).

Baltimore: Paul H. Brookes Publishing Co., 2001.

THE KEMPE FAMILY STRESS INVENTORY (KFSI)

Authors:

Barton Schmitt and Claudia Carroll with assistance from Jane Gray

Publisher:

Authors

www.kempecenter.org/about.htm

Initial Material Cost:

Free. The KFSI is not copyrighted and can be used at no charge and without the authors' permission, but permission is needed to re-publish the measure; supplemental rating criteria are copyrighted.

Representativeness of Norming Sample: None described.

Languages:

English

Type of Assessment:

Direct Parent Assessment

Age Range and Administration Interval: Not applicable. Focuses on parents of all ages.

Personnel, Training, Administration, and Scoring Requirements:

Personnel need to have appropriate training or experience to conduct psychosocial interview, and must have specialized training to use supplemental rating criteria.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Inter-rater: 3 (.65 or higher)

Validity: 2 (<.4 for predictive)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual).

Description: The KFSI assesses parents' risk for child maltreatment and/or caregiving difficulties. It is a 10-item rating scale that is completed after a through psychosocial interview with a trained professional. Service providers, such as home visitors, may also rate families based on the interactions they have had with them over a period of time (Korfmacher, Younge, and Michalek 1996). Items on the KFSI assess parents on a number of domains, such as psychiatric and criminal history, childhood history of care, emotional functioning, attitudes towards and perception of children, discipline of children, and level

of stress in the parent's life. Parents receive a raw score and are determined to be at low, moderate, or high risk, depending on the cut-offs established by the program administering the scale (see section on interpretability, below).

Uses of Information: The KFSI is currently used with at-risk families as an integral part of the screening and assessment process for two home-visiting programs for families at-risk, the Hawaii Healthy Start and the nation-wide Healthy Families America program. It is primarily used as a second-level screening tool for moth-

ers who are considered at risk based on a 15-item hospital chart review that assesses sociodemographic risk factors, such as maternal age and income.

Reliability: (1) Inter-rater reliability: a Healthy Families America program site in Oregon collected data on 115 families, and notes from each of their psychosocial interviews were reviewed by independent raters. The reliability correlation coefficient between raters on classifying parents into the low, mild, or severe risk categories was .93. (Katzev et al. 1997).

Validity: Predictive validity: three studies that compared KFSI scores to child maltreatment based on hospital records, state child protection reports, and scores on the Child Abuse Potential Inventory have shown KFSI to have predictive validity (Murphy et al. 1985; Hawaii Department of Health 1992; and Katzev et al. 1997). Three studies of predictive validity were conducted using varying cut-off scores and time periods. Positive predictive validity reflected the percentage of mothers with high-risk scores who later maltreated their children, and ranged from 25 to 52 in two of the studies. According to Korfmacher (2000), the low percentage (3 percent) in a third study may be an artifact of the study's reliance on child protection reports to measure the incidence of maltreatment. Negative predictive validity measured the percentage of mothers with low-risk scores who demonstrated no future evidence of child maltreatment; it ranged from 85 to 100 percent. The KFSI sensitivity, which is the percentage of maltreating mothers who were scored at high

risk, ranged from 80 percent to 97 percent across the three studies. The specificity, which is the percentage of non-maltreating mothers who were scored at low risk, ranged from 21 to 89 percent.

Method of Scoring: A three-point scale, ranging from low to high risk, is used to score both parents. Parents who receive a high-risk score are considered at risk for child maltreatment. Two methods have been employed in scoring the KFSI. Carroll (1978) used weighted scoring, in which items that were determined to be more immediate precedents to child abuse (such as violent outbursts and harsh punishment of child) were given higher scores, while lower weighting was given to items that were perceived to be less important in assessing immediate risk (such as parents' history of child maltreatment, low self-esteem or isolation, and having an unwanted child). If a weighted scaling system is not used, items are assigned values according to whether there is no risk (0), risk (5), or high risk (10) of child maltreatment. The total score is obtained by summing the values assigned to each item, although the weighted system requires that examiners multiply raw scores by 2.5. Total scores range from 0 to 100.

Interpretability: KFSI users have applied different cutoff scores based on their clinical judgment to identify at-risk parents. Carroll (1978) felt that scores above 30 were "concerning" and that scores above 70 were "particularly differentiating"; however, others have defined different cutoffs and have assigned labels (such as, low, medium and severe risk) to scores in a particular range. To aid in interpreting scores, Healthy

Families America and the Family Stress Center have created and copyrighted supplemental rating criteria.

Training Support: Individuals must participate in a specialized training on using the supplemental rating criteria, which have been copyrighted by Healthy Families America and the Family Stress Center.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References

Carroll, C.A. "The social worker's evaluation." In *The child protection team handbook*, edited by B.D. Schmitt. New York: Garland STM Press, 1978, pp. 83-108.

Center on Child Abuse Prevention Research. *Intensive home visitation: A randomized trial, follow-up, and risk assessment study of Hawaii's Healthy Start program*. Final report prepared for the National Center on Child Abuse and Neglect. Chicago, IL: National Committee to Prevent Child Abuse, 1996.

Hawaii Department of Health. Report to the

16th Legislature, State of Hawaii, on House Bill #139, c.d. 1: Requesting review and recommendations from the Director of Health on the Healthy Start Program. Honolulu, HI: Maternal and Child Health Branch, Hawaii Department of Health, 1992.

Katzev, A., T. Henderson, and C. Pratt. *Predicting child maltreatment with the Kempe Family Stress Assessment*. Document presented at HFA: Rethinking the Assessment Process working meeting. Chicago, IL: National Committee to Prevent Child Abuse, 1997.

Korfmacher, J., P. Younge, and P. Michalek. *Standardized assessment of child abuse risk: A review*. Paper presented at 24th Annual Child Abuse and Neglect Symposium, Keystone, CO, 1996.

Korfmacher, Jon. *The Kempe Family Stress Inventory: A review*. *Child Abuse and Neglect*, vol. 24, 2000, pp. 129-140.

Murphy, S., B. Orkow, and R.M. Nicola. *Prenatal prediction of child abuse and neglect: A prospective study*. *Child Abuse & Neglect*, vol. 9, 1985, pp. 225-235.

KNOWLEDGE OF INFANT DEVELOPMENT INVENTORY (KIDI), 1981

Authors:

David MacPhee

Publisher:

Unpublished manuscript; available from Educational Testing Service

(609) 734-5689

www.ets.org/

Initial Material Cost:

As of January 1998, the cost was \$11, plus \$3 shipping and handling to order this measure from the Educational Testing Service

Representativeness of Norming Sample: Non-representative sample of pediatricians, Ph.D.s in child psychology, University of North Carolina undergraduate child psychology students, and mothers in Chapel Hill, NC.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Parents of infants and young children

Personnel, Training, Administration, and Scoring Requirements:

An individual with a 7th-grade reading level can complete the instrument. Administration time is 20 minutes; scoring time is 20 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 2 (some less than .65; others .65 or higher)

Validity: 1 (validity coefficients not reported)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

Description: The Knowledge of Infant Development Inventory (KIDI) is a 75-item instrument that was designed to obtain comprehensive information on parents' factual knowledge of parental practices, child developmental processes, and infant norms of behavior. The KIDI is designed to be easily accessible to persons with limited education and to be culturally neutral. The items can also be grouped into four non-exclusive general categories to obtain more specific information on a person's knowledge on

infant norms and milestones, principles of infant development, parenting, and health and safety. The KIDI Scale is accompanied by a 17-item questionnaire (the Catalog of Previous Experience, or COPE) assessing previous experience with infants to correlate with knowledge level assessed by KIDI.

Uses of Information: The KIDI may be used as an indicator or a diagnostic tool for high-risk parents and also to evaluate parent education programs.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): alphas were .67 and .55 for college students at pretest and posttest, respectively, .82 for parents, and .50 for professionals. The Guttman split-half coefficients were .60 and .57 for college students at pretest and posttest, respectively, .85 for mothers, and .59 for professionals. (2) Test-retest reliability: For parents (2-week interval), the correlation coefficients were .92 for the total score, .80 for attempted, and .91 for accuracy. For college students (4-month interval), the coefficients were .65 for the attempted and .47 for accuracy.

Validity: (1) Content validity: The author conducted an extensive review of the relevant literature and the instrument has been reviewed by parents, pediatricians, and persons holding a Ph.D. in child psychology. (2) Construct validity: The manual reports the results of the initial validity studies conducted by the author. The results suggested that persons with more experience with or knowledge about infants were more confident in responding to the KIDI. However, persons with formal knowledge were more accurate in their responses than persons with informal knowledge. (3) Predictive validity: In another study, the author found parents of developmentally delayed children had significantly lower KIDI attempted and accuracy scores than parents of children with normal development.

Method of Scoring: Each of the KIDI items is scored as right (+1), wrong (-1), or not sure (0) according to an answer key that is provided. Using formulas provided in the manual, three

summary scores are then calculated: an attempted score (percent of items attempted, a measure of confidence), an accuracy score (percent correct of the attempted answers), and a total correct score (percent correct of all the KIDI items). If the user wishes, subscale scores can be calculated for the four general categories: (1) norms and milestone, (2) principles, (3) parenting, and (4) health and safety.

Interpretability: No instructions provided.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

MacPhee, D. Manual: Knowledge of Infant Development Inventory. Unpublished manuscript, University of North Carolina, 1981.

MacPhee, D. The Nature of Parents' Experiences with and Knowledge About Infant Development. Paper presented at the Society for Research in Child Development, April 1983.

MacPhee, D. The Relationship Between Children's Delayed Development and Their Mothers' Perceptions of Development. Paper presented at the Southwestern Society for Research in Human Development, March 1984.

MacPhee, D. Mothers' Acquisition and Reconstruction of Knowledge About Infancy. Paper presented at the Southwestern Society for Research in Human Development, March 1984.

NURSING CHILD ASSESSMENT SATELLITE TRAINING (NCAST) PARENT-CHILD INTERACTION PROGRAM NURSING CHILD ASSESSMENT TEACHING SCALE (NCATS), 2ND EDITION, 1995

Authors:

NCAST

Publisher:

NCAST

(206) 543-8528

www.ncast.org

Initial Material Cost:

Teaching set (includes teaching manual, scale pad, and teaching kit): \$125

Representativeness of Norming Sample: Diverse but non-random sample. Sample consists of approximately 2,100 observations sent in by persons across the United States learning to use the scale from around 1980 to 1995.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval: Birth to 36 months

Personnel, Training, Administration, and Scoring

Requirements: Administered by a professional health care worker certified by NCAST as a learner or instructor. Usually administered in 1 to 6 minutes.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (some subscales fell below .65, although most of the total scales exceeded .65)

Validity: 2 (<.5 for concurrent, <.4 for predictive)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Nursing Child Assessment Teaching Scale (NCATS) is used to assess the quality of the caregiver-child teaching interaction for children from birth to 3 years of age. The 73-item teaching scale is organized into six subscales, four of which assess the caregiver's behavior and two the child's. The four caregiver subscales assess the caregiver's sensitivity to cues, response to the child's distress, fostering of social-emotional growth, and fostering of cognitive growth. The two child subscales assess the clarity of the child's cues and responsiveness to the caregiver.

Uses of Information: The teaching scale identifies areas of strengths and weaknesses in the

caregiver-child teaching interaction. The results can be used to build the caregiver's skills to facilitate the development of the child.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): ranged from .52 to .80 on the caregiver subscales, .50 on the child's clarity of cues, and .78 on the child's responsiveness to parent subscales. The alphas for the total caregiver and child subscales were .87 and .81, respectively. (2) Test-retest reliability (with a 3- to 4-month interval between tests): .85 on the total parent score and .55 on the total infant score.

Validity: (1) Concurrent validity: NCATS

caregiver scores were tested for concurrent validity against the Home Observation for Measurement of the Environment (HOME) and the Bayley Scales of Infant Development. The correlations of the total NCATS scores with the total HOME score among children ages 1 to 36 months, in three age groups, ranged from .41 to .44. The correlation of the total NCATS score with the Bayley Mental Development Index (MDI) and Bayley Psychomotor Development Index (PDI) were .28 and .34, respectively. In both cases, the caregiver scales, especially the social-emotional and cognitive growth subscales, were more strongly correlated with the HOME and Bayley. (2) Predictive validity: a test for predictive validity reported correlations of .23 and .34, respectively, between NCATS total scores taken at 3 and 10 months and MDI scores, both statistically significant. The subscale correlations ranged from $-.01$ to .37. Correlations between the NCATS caregiver and total scores at 24 months with the Bayley MDI (at 24 months), Preschool Language (at 36 months), and WPPSI IQ (at 60 months) were stronger and more consistent than the correlations between the cognitive measures and NCATS scores at 12 months.

Method of Scoring: During the teaching session, the observer goes through the 73-item scale and marks “yes” or “no” for each item depending upon whether or not the behavior was observed. The teaching manual provides the user with step-by-step scoring instructions. The user must calculate the totals for each subscale and the total score. The scores are compared to a table provid-

ed in the manual to determine whether the score falls under the 10 percentile cutoff score.

Interpretability: The Teaching Manual provides step-by-step instructions on how to interpret scores using population norms. Tables that compare NCATS scores to the norms are provided.

Training Support: To learn essential child care skills, users of NCATS are strongly recommended to view NCAST’s “Keys to Caregiving” video series. Workshops are also available through NCAST or NCAST certified instructors. The fee for NCAST’s workshops is \$900, which covers training on assessing caregiver-child interaction. NCAST recommends that individuals view the “Keys to Caregiving” videos prior to attending NCAST training.

Adaptations/Special Instructions for Individuals with Disabilities: None.

Report Preparation Support: General guidance is provided in the step-by-step instructions on the use of the teaching scale, including some suggestions on how to review and discuss any identified problems with caregivers and how to document an agreed upon prescription of practice/behavior by the caregivers.

References:

Summer, Georgina and Anita L. Spietz. NCAST Caregiver/Parent-Child Interaction Teaching Manual, 2nd Edition. Seattle, WA: NCAST Publications, University of Washington, June 1995.

PARENTING STRESS INDEX, THIRD EDITION (PSI), 1995

Authors:

Richard R. Abidin

Publisher:

Psychological Assessment Resources, Inc.

(800) 331-8378

www.parinc.com

Initial Material Cost:

PSI Long Form Kit: \$131 (includes manual, 10 reusable item booklets, 25 hand-scorable answer sheet/profile forms)

PSI Short Form Kit: \$90 (includes manual, 25 hand-scorable questionnaire/profile forms)

Representativeness of Norming Sample: The English version of the PSI was standardized with parents of children ranging from 1 month to 12 years (mean of 4.9). The non-random sample of parents included 2,633 mothers (ages ranging from 16 to 61, with a mean of 30.9) and 200 fathers (ages ranging from 18 to 65, with a mean of 32.1). The parents were recruited by clinic, school, or child care center staff and volunteered to participate in the norm sampling study. The Spanish version was normed on a sample of 223 Hispanic parents.

Languages:

English, Spanish, French

Type of Assessment:

Parent report

Age Range and Administration Interval: For parents of children ages 1 month to 12 years

Personnel, Training, Administration, and Scoring Requirements:

The manual states that an individual without formal training in psychology or social work can administer and score the PSI, but the interpretation of PSI scores requires someone with training in these or other related disciplines. Parent needs to have at least a 5th grade education. The PSI long form takes approximately 20 to 30 minutes to complete.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (internal consistency and test-retest .65 or higher for both the Long and Short Forms)

Validity: 3 (mostly .5 or higher for concurrent validity)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (no special administration requirements, scored by highly trained individual)

Description: The purpose of the 120-item PSI is to produce a diagnostic profile of perceived child and parent stress. The PSI was developed based on the theory that total parental stress is a function of child and parent characteristics, as well as situational variables. It contains 13 subscales within 4 major domains: total stress, child

domain, parent domain, and life stress. The total stress domain, which measures the level of stress in the parent-child relationship, is comprised of the child and parent domains. The child domain has six subscales that measure the child's distractibility/hyperactivity, adaptability, reinforcement of the parenting experience, demanding-

ness, mood, and acceptability. The remaining seven subscales make up the parent domain and measures: competence, isolation, attachment, health, feeling of role restriction, depression, and spousal support. The life stress domain measure sources of stress beyond the parent's control. The PSI is also available in a Short Form, which consists of a 36-item self-scoring questionnaire and profile.

Uses of Information: Primary uses are screening for early identification, assessment for individual diagnosis (including informing therapy and counseling), pre-post measurement for effectiveness of intervention, and research for studying the effects of stress on parent-child interactions.

Reliability: (1) Internal consistency (Cronbach's alpha) for the PSI (Long Form) subscales ranged from .70 to .83 in the Child Domain, .70 to .84 in the Parent Domain, and was greater than .90 for the two domains and the Total Stress scale. Similar internal consistency alphas for the PSI were also established in a cross-cultural population study (Hauenstein, et al., 1987). In the PSI Short Form (PSI/SF) subscales, internal consistency (Cronbach's alpha) was .85 in the Difficult Child, .80 in Parent-Child Dysfunctional Interaction, .87 in Parental Distress, and .91 in Total Stress. (2) Test-retest reliabilities (intervals between administrations of the PSI in these studies ranged from 3 weeks to 1 year) in the PSI Long Form ranged from .55 to .82 for the Child Domain, .69 to .91 for the Parent Domain, and .65 to .96 for the Total Stress score.

Validity: (1) Concurrent validity: the manual provides an abstract of studies that demonstrated concurrent validity by comparing PSI (Long Form) scores with those on other assessment instruments. Only a few of the abstracts provided validity statistics. The few that reported statistics found that the correlation between Total Stress and the Bayley Scale was .42 at 3 months and .66 at 6 months. The correlation between child domain and negative behavior in hyperactive siblings relationships was .60, while its correlation with the 6 factors in the Family Impact Questionnaire ranged from .36 to .84. A study also reported correlations ranging from .65 to .77 between life stress and the lack of formal support among parents of children with disabilities.

Method of Scoring: The PSI contains a hand-scorable Answer Sheet on which basic demographic information and item responses are included. Most responses require the respondent to circle SA (strongly agree), A (agree), NS (not sure), D (disagree), or SD (strongly disagree) in response to the particular items. Addition and, if there are missing data, division skills are needed to obtain the raw scores. Using the profile form, which is part of the answering sheet, the scorer can obtain the percentile ranking for each subscale score. The respondent's score can also be graphed on the profile form. Detailed information on scoring is provided in the Professional Manual.

The PSI also offers a Software Portfolio, Windows software that allows you to administer either the 120-item PSI or the 36-item PSI Short

Form on-screen or to enter item responses from the PSI or the PSI Short Form. The software automatically scores the item responses and generates a report. Reports can be edited on-screen. This updated software program contains modifications to the interpretive statements, empirically based cutoff scores, and reference lists of PSI research.

Interpretability: The manual states that interpretation of the PSI scores requires someone who has graduate training in clinical, counseling, or educational psychology or in social work or a related field. Interpretation guidelines are discussed in the manual, and it is suggested that the individual reviewing and interpreting the results first interpret the Total Stress score, and then look at the Child Domain and Parent Domain scores and their subdomains scores to pinpoint the sources of stress. Throughout the interpretation guidelines in the manual, there are references to research literature. The interpretation section also includes five case illustrations profiling different parental and situational characteristics.

Training Support: None mentioned in manual

Adaptations/Special Instructions for Individuals with Disabilities: While there are no

explicit instructions for administering the PSI with parents of children with disabilities, the manual contains information on how the PSI may work with this population. Sections entitled “Families with Special-Needs Children” and “Disabilities and Illnesses” cites various research studies related to use of the PSI in families with children having some disabilities. These studies are summarized, and cover various disabilities including: autism, deafness, congenital heart disease, asthma, cystic fibrosis, and so forth.

Report Preparation Support: The software generates a report. Two sample reports can be found at the publisher’s website: www.parinc.com

References:

Abidin, Richard R. Parenting Stress Index, Third Edition. Odessa, FL: Psychological Assessment Resources, 1995.

Hauenstein, E., S. Scarr, and Richard Abidin. Detecting Children At-risk for Developmental Delay: Efficacy of the Parenting Stress Index in a Non-American Culture. Unpublished manuscript. Charlottesville: University of Virginia, 1987.

SUPPORT FUNCTIONS SCALE (SFS), 1985

Authors:

Carl J. Dunst and Carol M. Trivette

Publisher:

Brookline Books

(617) 558-8010 or (800) 666-BOOK

www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling And Empowering Families, \$25

\$10 per batch of 10 score sheets.

Representativeness of Norming Sample: No norming sample. The research sample consisted of 121 parents of preschool mentally challenged, handicapped, and developmentally at-risk children.

Languages:

English

Type of Assessment:

Parent self-report

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent (or caregiver) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 2 (.65 or higher for internal, split-half, and test-retest reliability for individual items; <.65 for test-retest reliability for total scale score)

Validity: 2 (<.5 for criterion validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The self-report Support Functions Scale (SFS) measures the extent of parents' needs for different types of support. The scale is available in both an extended (20-item) and short (12-item) version. Both versions ask parents to rate their need for financial, emotional, instrumental, and informational support on a five-point scale ranging from never (1) to quite often (5).

Uses of Information: Providers can use the SFS results to guide follow-up discussions with

parents to better understand their needs and to develop an intervention plan to address needs. The book provides a number of case studies to illustrate how the information can be used.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): .87; split-half reliability (using the Spearman-Brown formula): .88; (2) Test-retest reliability (1-month interval): the average correlation among administrations for the individual items was .91; for the total scale score, the correlation among administrations was .62.

Validity: (1) Concurrent validity: The authors compared the outcomes on the SFS scale to the outcomes on scales measuring family well-being (McCubbin et al.), personal well-being (Trivette and Dunst, 1985), and time demand on respondent (Dunst and Trivette, 1985). The total scores (20-item scale) proved to be the best predictor. Both family (correlation = .27) and personal (correlation = .33) well-being were significantly related to adequacy of support, whereas lack of support placed more time demands upon the respondent (correlation = -.20). Financial support was the only factor score significantly related to family well-being (correlation = .27), whereas emotion (correlation = .17), child-related (correlation = .21), and instrumental (correlation = .29) support were significantly related to personal well-being. None of the factor scores were related to the personal time demands measure.

Method of Scoring: The parent or caregiver answers to what extent he or she feels a need for each type of assistance by marking Never, Once in a While, Sometimes, Often, and Quite Often.

Interpretability: If the respondent rates an item as Sometimes, Often, or Quite Often (have a need), this may be taken as an indication that further interviewing (assessment) is necessary to determine the exact type of help that is needed but lacking.

Training Support: None. However, a second book published by Brookline Books, entitled "Supporting & Strengthening Families: Methods Strategies and Practices" is a collection of papers updating the thinking and practices described in

"Enabling and Empowering Families", and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None, this measure was developed based on work with families with children who have disabilities.

Report Preparation Support: The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to identify providers and record the resources they are expected to provide to help the family address an identified need. The family support form helps the agency to mobilize resources to address needs and to monitor the progress of the intervention.

References:

Dunst, Carl and Carol Trivette. Personal Time Commitment Scale: Reliability and Validity. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program, 1985.

Dunst, Carl, Carol Trivette, and Angela Deal. Enabling and Empowering Families: Principles & Guidelines for Practice. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. Supporting and Strengthening Families: Methods, Strategies, and Practices. Newton: Brookline Books, 1994.

McCubbin, H.I., J.K. Comeau, and J.A. Harkins. "Family Inventory of Resources for Management." In H.I. McCubbin and J.M.

Patterson (eds.), Systematic Assessment of Family Stress, Resources, and Coping. St. Paul, MN: Family Stress and Coping Project, 1981, pp. 67-69.

Trivette, Carol and Carl Dunst. Personal

Well-Being Index: Reliability and Validity. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1985.

Program Implementation and Quality Instruments^a



Program Implementation and Quality Instruments^a

Instrument	Domains	Age Range	Assessment Type	Initial Material		Ease of Administration		
				Cost	Reliability	Validity	Norming Sample	Scoring
Arnett Caregiver Interaction Scale	Child care quality	0 - 5.5 years	2	1	3	3	1	3
Early Childhood Environment Rating Scale- Revised	Child care quality	2.5-5 years	2	1	3	1	1	3
Early Head Start Evaluation--Parent Services Interviews	Program services	NA	3	1	1	1	1	2
Family Day Care Rating Scale	Child care quality	0-5 years	2	1	3	1	1	3
Head Start Family Information System ^b	Program services, including home visit documentation	NA	3	1	1	1	1	3
Infant/Toddler Environment Rating Scale	Child care quality	0-2.5 years	2	1	3	3	1	3
National Association for the Education of Young Children Accreditation Criteria	Child care quality	NA	3	1	3	NA	NA	1
Program Implementation Checklist and Rating Scales	Program quality	NA	3	1	1	NA	NA	3
Program Review Instrument for Systems Monitoring	Program quality	NA	2,3	1	1	1	1	1

KEY

Assessment Type	Initial Material Cost	Reliability	Validity
1 = Direct assessment	1 = Under \$100	1 = None described	1 = None described
2 = Observation	2 = \$100 to \$200	2 = Under .65	2 = Under .5 for concurrent; under .4 for predictive
3 = Parent/self report	3 = More than \$200	3 = .65 or higher	3 = .5 or higher for concurrent; .4 or higher for predictive

Norming sample

- 1 = None described
- 2 = Older than 15 years, not nationally representative or representative of EHS population
- 3 = Normed within past 15 years, nationally representative or representative of EHS population

Ease of administration and scoring

- 1 = Not described
- 2 = Self-administered or administered and scored by someone with basic clerical skills
- 3 = Administered and scored by a highly trained individual

^aThe information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

^bThe HSFIS is a management information system. It also includes measures of home-based service use and forms for tracking child and family outcomes.

ARNETT CAREGIVER INTERACTION SCALE, 1989

Authors:

Jeffery Arnett

Publisher:

None.

A copy of the scale can be found in Jaeger and Funk (2001)

Cost:

None

Representativeness of Norming Sample: None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval: Caregivers of early childhood classes

Personnel, Training, Administration, and Scoring Requirements:

To be a certified Arnett Caregiver Interaction Scale observer requires achieving a .70 inter-rater reliability coefficient for two consecutive visits. (Jaeger and Funk). No recommended length of observation. Arnett observed caregivers in two 45-minute sessions, while Jaeger and Funk observed caregivers in a 2.5-hour session.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: Internal consistency and inter-rater reliability: 3 (.65 or higher)

Validity: Concurrent: 3 (mostly .5 or higher)

Norming Sample Characteristics: 1 (none described)

Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The 26-item Caregiver Interaction Scale assesses the quality and content of the teacher's interactions with children. The scale was designed to provide information on various socialization practices that have been identified in research on parenting. The scale can be used without modification in both center and

home-based settings. The items measure the emotional tone, discipline style, and responsiveness of the caregiver in the classroom. The items are usually organized into the following four subscales: (1) positive interaction (warm, enthusiastic, and developmentally appropriate behavior), (2) punitiveness (hostility, harshness, and use of

¹ The scale is also referred to as the Arnett Scale of Caregiver Behavior.

threat), (3) detachment (uninvolvement and disinterest), and (4) permissiveness.

Uses of Information: The scale can be used to assess caregiver's interactions with children and their emotional tone and approach to engaging and disciplining children.

Reliability: (1) Internal consistency: Layzer et al. obtained Cronbach alphas of .91 for warmth/responsiveness (positive interaction) and .90 for harshness (punitiveness), while Resnick and Zill obtained alphas for the total scale of .98 for lead teachers and .93 for assistant teachers. Jaeger and Funk reported coefficients of .81 and higher for the sensitivity (positive interaction), punitiveness, and detachment subscales. (2) Inter-rater reliability: Jaeger and Funk reported inter-rater reliability coefficients ranging from .75 to .97 between a certified observer and trainees.

Validity: (1) Concurrent validity: Layzer et al. reported correlation coefficients of .43 to .67 between the Arnett and the Early Childhood Environment Rating Scale (ECERS), Assessment Profile for Early Childhood Programs, and the Description of Preschool Practices. The authors did not expect the coefficients to be large because the Arnett scale focused more narrowly on an aspect of teacher behavior not directly measured by the other three observation instruments. However, Phillipsen et al. reported a correlation of .76 between the Arnett and the ECERS.

Method of Scoring: The observer rates the extent to which the caregiver exhibits the behavior described in the item on a 4-point scale, ranging from not at all (1) to very much (4). Averages can

be calculated for each subscale.

Interpretability: Depending on the program's needs, individual caregiver scores can be compared to the scores of other caregivers or the mean scores of a group of caregivers compared against the means of other groups of caregivers. Statistical tests have been frequently utilized to assess the differences between scores.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Arnett, Jeffery. "Caregivers in Day-Care Centers: Does Training Matter?" *Journal of Applied Developmental Psychology*. Vol. 10, 1989, pp. 541-552.

Jaeger, Elizabeth, and Suzanne Funk. *The Philadelphia Child Care Quality Study: An Examination of Quality in Selected Early Education and Care Settings*. Philadelphia: Saint Joseph's University, October 2001.

Layzer, Jean I., Barbara D. Goodson, and Marc Moss. *Observational Study of Early Childhood Programs, Final Report, Volume I: Life in Preschool*. Cambridge, MA: Abt Associates, Inc., 1993.

Phillipsen, Leslie, Debby Cryer, and Carollee Howes. "Classroom Process and Classroom Structure." In *Cost, Quality, and Child Outcomes in Child Care Centers*, edited by Suzanne W. Helburn. Denver: Department of Economics,

Center for Research in Economics and Social Policy, University of Colorado at Denver, 1995, pp. 125-158.

Resnick, Gary, and Nicholas Zill. Is Head Start Providing High-Quality Education Services? "Unpacking" Classroom Processes. Albuquerque, NM: Biennial Meeting of the Society for Research in Child Development, April 15-18, 1999.

U.S. Department of Education. National Center for Education Statistics. Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research, Working Paper No. 97-36, by John M. Love, Alicia Meckstroth, and Susan Sprachman. Jerry West, Project Officer. Washington, DC: 1997.

EARLY CHILDHOOD ENVIRONMENT RATING SCALE - Revised Edition (ECERS-R), 1998

Authors:

Thelma Harms, Richard M. Clifford, and Debby Cryer
www.fpg.unc.edu

Publisher:

Teachers College Press
www.teacherscollegepress.com
1-800-575-6566

Initial Material Cost:

ECERS-R Assessment Scale: \$13; Video training: \$59;
Workbook: \$4

Representativeness of Norming Sample: Research reported in manual does not include this information

Languages:

English¹

Type of Assessment:

Observation, with some caregiver report

Age Range and Administration Interval: For classrooms enrolling children 2 1/2 to 5 years of age.

Personnel, Training, Administration, and Scoring Requirements:

The individual administering the ECERS-R must read and practice the scale, and must also have knowledge of child development and educational implications. The authors recommend reviewers to have at least two practice observation sessions with an experienced ECERS-R trainer. Administration time ranges from 2 hours to 5 hours depending on the scoring option selected for administering the assessment.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher for internal consistency)

Validity: 1 (concurrent not available, predictive is promising)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The ECERS-R is designed to assess group programs for children of preschool age (2 1/2 to 5). It is a 43-item assessment tool rating scale organized into seven environmental

subscales: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. Each item has a number of quality indica-

¹ Please contact Thelma Harms to obtain information about other official translations

tors. The ECERS-R can be used in preschool, kindergarten, and child care classrooms. The original ECERS was revised to reflect changes in the early childhood field and to be more inclusive of children with disabilities and sensitive to cultural diversity.

Uses of Information: The assessment can be used by program directors for supervision and program improvement, by teaching staff for self-assessment, by agency staff for monitoring, and in teacher training programs.

Reliability: (1) Internal consistency (Cronbach's alpha) ranged from .71 to .88 at the subscale level; the total scale internal consistency was .92. (2) Inter-observer reliability was .92 (Pearson correlation) and .87 (Spearman correlation).

Validity: Concurrent validity information is not available in the manual. However, the original version of the ECERS has been demonstrated to have good predictive validity (Peisner-Feinberg and Burchinal 1997), indicating that quality of center-based child care (measured using the ECERS-R) was related to preschool children's concurrent cognitive and socioemotional development (measured using the Peabody Picture Vocabulary Test-Revised and the Woodcock-Johnson Tests of Achievement-Revised).

Method of Scoring: The scoring sheet records the ratings for quality indicators, items, subscale and total scores, as well as any observer comments. The indicators, which have Yes/No/Not applicable response choices, are used to score the items, which have scores ranging from 1

(Inadequate) to 7 (Excellent). There are two ways to score the items. The manual provides detailed instructions on both scoring systems. The usual scoring system for each item is based on the number of the highest quality indicators with affirmative responses. Under the alternate scoring method, each indicator is individually scored (using the 1 to 7 range), which could extend the assessment time to a total of 4 to 5 hours. A Profile sheet is also provided to graphically display the scoring information, to compare areas of strengths and weaknesses, and to select items and subscales to target for improvement. The profiles for at least two observations can be plotted side by side to depict changes visually. Sample profile forms are available on the Frank Porter Graham Child Development Center website (www.fpg.unc.edu).

Interpretability: Full instructions for using the scale, plus notes clarifying selected scale items are included. However, the manual provides no information about interpreting the results of the observation.

Training Support: The assessment tool includes instructions for using the ECERS-R; a Video training package and workbook are also available. The website: www.fpg.unc.edu includes helpful information.

Adaptations/Special Instructions for Individuals with Disabilities: The revised version includes new items to assess program's services for children with disabilities. However, the revision needs field-testing and standardization.

Report Preparation Support: Not included in the manual.

References:

Harms, Thelma, Richard M. Clifford and Debby Cryer. Early Childhood Environment Rating Scale, Revised Edition. New York: Teachers College Press, 1998.

Harms, Thelma and Debby Cryer. Early Childhood Environment Rating Scale Video Observations, Revised Edition. New York: Teachers College Press, 1999.

Harms, Thelma and Debby Cryer. Early Childhood Environment Rating Scale Video Guide & Training Workbook, Revised Edition. New York: Teachers College Press, 1999.

Peisner-Feinberg, E., and M. Burchinal. "Relations Between Preschool Children's Child Care Experiences and Concurrent Development: The Cost, Quality and Outcomes Study." Merrill-Palmer Quarterly, vol. 43, no. 3, 1997, pp. 451-477.

www.fpg.unc.edu

NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT PARENT SERVICES INTERVIEWS, 1996 - 2001

Authors:

John Love and other project staff, in collaboration with the Early Head Start Research Consortium

Publisher:

Mathematica Policy Research, Inc.
Contact Publications, 609-275-2350,
jallen@mathematica-mpr.com. The interviews can be found at www.mathematica-mpr.com or www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_instruments.html.

Initial Material Cost:

None

Representativeness of Norming Sample: None described

Languages:

English, Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval:

For child-related questions, prenatal through 36 months.

For parent-related questions, all adults.

Personnel, Training, Administration, and Scoring Requirements:

The questions were designed to be administered to parents by someone with basic interviewing skills. Very little scoring is required.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered by someone with basic interviewing skills)

Description: The Parent Services Interviews (PSIs) developed for the national Early Head Start Research and Evaluation Project were designed to include instruments that assess potential program effects on service needs and use, as well as outcomes related to economic self-sufficiency. The interview questions were drawn from a variety of sources and include published instruments, questions drawn from other large national surveys, and questions that were developed specifically for this study. They cover topics including: family goals, perceived needs and resources, employment, education and job training, child care,

home visits, transportation, housing, social support, health status, health care services, case management, other family support services, and public assistance receipt. The interviews can be found on the web at the addresses listed above.

The results of the evaluation (through age 3) are included in two reports and their appendices and are available at www.mathematica-mpr.com and www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_instruments.html.

Uses of Information: The Early Head Start

PSIs can be used by programs to obtain service use information that can be compared to the national study findings. The summary information on services obtained from parents in the PSI questions may be especially useful to programs that do not collect detailed service use information in a management information system. For copyrighted instruments, programs must obtain permission to use the assessments and must pay for their use.

Reliability: The technical appendices of the two reports include internal consistency reliability for summary scores. As a general rule, summary scores were not included in the report if their reliability was not above .65. Most measures in the PSI interviews were single questions and did not require computing summary scores.

Validity: Many of the questions were included in the evaluation because they had been used before in large studies and had demonstrated construct validity. Validity work based on the data collected was not reported in the two reports.

Method of Scoring: Most PSI measures were based on single questions and do not need scoring. Scoring procedures for any measure requiring scoring are summarized in the reports.

Interpretability: The information obtained from the PSI questions is easily interpretable.

Training Support: As part of the evaluation project, in-depth training manuals were developed; these can be obtained by requesting them from Jackie Allen at jallen@mathematica-mpr.com. In addition to the manuals, interviewers attended a central training session and had to meet rigorous standards before administering the

study instruments. Mathematica is not providing any training support for the measures.

Adaptations/Special Instructions for Individuals with Disabilities: Contact Mathematica for more information about how the protocols were adapted for use with individuals with disabilities.

Report Preparation Support: None described.

References:

ACYF. "Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report." Washington, DC: Administration on Children, Youth and Families, 2001. www.mathematica-mpr.com or http://www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_reports.html.

ACF. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC: Administration for Children and Families, June 2002. www.mathematica-mpr.com or http://www.acf.dhhs.gov/programs/core/ongoing_research/ehs/ehs_reports.html.

Sprachman, S., C. DeSaw, L. Mendenko, M. Salem, K. Boller, and B. Kolln. "Early Head Start National Evaluation Data Collection Training Manual for 6-Month Parents Services Interview, 14-Month Parent-Child Interview and Assessment, and Child Care Observations." Princeton, NJ: Mathematica Policy Research, Inc., December 1996.

www.mathematica-mpr.com

FAMILY DAY CARE RATING SCALE (FDCRS), 1989

Authors:

Thelma Harms and Richard M. Clifford
www.fpg.unc.edu

Publisher:

Teachers College Press
1-800-575-6566
www.teacherscollegepress.com

Initial Material Cost:

FDCRS Assessment: \$13
Video Observations: \$59
Video Guide and Training Workbook: \$4
30 Scoring Sheets: \$9

Representativeness of Norming Sample: None described

Languages:

English¹

Type of Assessment:

Observation and self-assessment by program staff

Age Range and Administration Interval: From 0 - 5 years

Personnel, Training, Administration, and Scoring Requirements:

Possible for supervisor, researcher, or trained day care provider to administer the FDCRS. Prior to administration, approximately 2 hours of reviewing the scale, 1 to 2 hours of video training, and two practice classroom observations (2 hours each) are recommended. FDCRS observation takes approximately 2 hours.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher)

Validity: 1 (concurrent not available, predictive is promising)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The FDCRS is an adaptation of the Early Childhood Environment Rating Scale - ECERS (see ECERS-R review in this section), but focuses on the quality of family day care settings rather than center-based settings. The scale can be used by a supervisor, researcher, or trainer

during an observation, or by a care provider as a self-assessment. The FDCRS contains 32 items organized in 6 subscales: Space and Furnishings for Care and Learning, Basic Care, Language and Reasoning, Learning Activities, Social Development, and Adult Needs. Each item is

¹ The FDCRS has also been translated into French. Those interested may contact Thelma Harms at the address above.

rated from 1 to 7 with quality descriptors associated with levels: (1) inadequate (does not even meet custodial care needs); (3) minimal (meets custodial needs, and to some extent, basic developmental needs); (5) good (meets developmental needs); and (7) excellent (high-quality personalized care).

Uses of Information: The FDCRS was designed to be comprehensive, yet easy to use as part of supervision and monitoring by agency staff, self-evaluation by care providers, and also in research and program evaluation.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) for the subscales: Space and Furnishings for Care and Learning (.86), Basic Care (.90), Language and Reasoning (.90), Learning Activities (.93), Social Development (.83), and Adult Needs (.70); (2) Inter-rater reliability was .90 for individual items in studies reported by Howes and Stewart (1987) and Howes (1987). Reliability has been demonstrated for diverse groups in subsequent research.

Validity: (1) Concurrent validity: information not available in the manual. However, studies showed that scores on the earlier versions of the FDCRS were highly correlated (.80) with home visitors' ratings of family day care settings. The FDCRS is also related to caregiver education and child outcomes.

Method of Scoring: The packet includes a Score Sheet, to record scores for individual items. The Profile sheet permits a graphic representation

of the scoring information. It can be used to compare areas of strengths and weaknesses, and to select items and subscales to target for improvement. The profiles for at least two observations can be plotted side by side to depict changes in a family child care home over time or differences between settings visually.

Interpretability: Full instructions for using the scale, plus notes clarifying selected scale items are included. However, the manual provides no information about interpreting the results of the observation.

Training Support: Video Observations and a Video Guide and Training Workbook are available for an additional cost. This multimedia package demonstrates how to use the Family Day Care Rating Scale. Each training package contains an interactive videotape and an Instructor's Guide, which explains how to present the various training activities and provides answers and explanations for any questions that may arise. A 16-page Video Guide and Training Workbook contains training activities. The website: www.fpg.unc.edu also includes helpful information.

Adaptations/Special Instructions for Individuals with Disabilities: FDCRS includes Supplementary Items for Exceptional Children to be used when the facility enrolls a special needs child.

Report Preparation Support: Not included in the manual.

References:

Harms, Thelma, and Richard M. Clifford.
Family Day Care Rating Scale. New York:
Teachers College Press, 1989.

Harms, Thelma, and Richard M. Clifford.
Family Day Care Rating Scale—Video Guide and
Training Workbook. New York: Teachers College
Press, 1993.

Harms, Thelma, and Richard M. Clifford.
Family Day Care Rating Scale—Video
Observations. New York: Teachers College Press,
1993.

Howes, C. Inter-Observer Reliability for the
Harms and Clifford Family Day Care Rating Scale.
Personal Communication, 1987.

Howes, C. and P. Stewart. “Child’s Play with
Adults, Toys, and Peers: An Examination of
Family and Child Care Influences.”
Developmental Psychology, vol. 23, 1987, pp. 423-
430.

www.fpg.unc.edu

HEAD START FAMILY INFORMATION SYSTEM (HSFIS) 4.3, 2001

Authors:

Head Start Bureau
www.acf.hhs.gov/programs/hsb/programs

Publisher:

Cleverex Systems, Inc.
301-738-1122
www.cleverex.com
info@cleverex.com

Initial Material Cost:

HSFIS is available free of charge to Head Start programs.

Representativeness of Norming Sample: None described.

Languages:

English, Spanish

Type of Assessment:

Management information system, including measures of family needs, center-based services, and home-based services, as well as features for tracking child and family outcomes.

Age Range and Administration Interval:

For child-related information, prenatal to school age. For parent- and service-related questions, all adults.

Personnel, Training, Administration, and Scoring Requirements:

Varies. HSFIS is a management information system designed for use in tracking services and outcomes.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered by a trained individual)

Description: The Head Start Family Information System (HSFIS) is an automated case management record-keeping system. It is designed to collect, organize, maintain, and report on information at both the child and family level to assist agencies in managing and improving services. The next generation HSFIS, a new web-based system called PROMIS (Program Resources and Outcomes Management Information System), has been developed and is being piloted in early 2003. It will include a

module that can be used to track the progress and accomplishments of children in efforts to analyze and use data on child outcomes in program self-assessment and continuous improvement. It will be based on the framework set forth in ACF-HS-IM-00-18, "Using Child Outcomes in Program Self-Assessment."

Uses of Information: Information collected in HSFIS can help programs determine eligibility for enrollment, identify family service needs, provide program-level demographic statistics, track

delivery of services to children and families, and track child and family outcomes.

Reliability: Not applicable.

Validity: Not applicable.

Method of Scoring: Not applicable.

Interpretability: Not applicable.

Training Support: Cleverex Systems, Inc. (<http://support.cleverex.com>, or 1-800-473-4780) provides training and technical assistance in the use of HSFIS and PROMIS. Cleverex Systems provides instruction at its own computer laboratory in Rockville, Maryland and at local and regional training events. Training courses are offered at three levels -- beginner, intermediate and advanced. Also, a 1-800 Help Desk provides technical assistance to users. Each Head Start Regional Office and Quality Improvement Center has a designated HSFIS liaison who can assist programs in obtaining automation funding and technical support.

Adaptations/Special Instructions for Individuals with Disabilities: None necessary.

Report Preparation Support: The HSFIS software includes the capability of producing more than 200 pre-defined reports. An Ad-Hoc Report Builder has been developed in the system so that users can develop their own reports without additional training. In addition, advanced users can use Crystal Report software to develop more-sophisticated reports using HSFIS data.

References:

Angulo, Paulo. HSFIS Update. www.cleverex.com/ShowArticle.asp?id=HSFIS1. February 2000.

Gaidurgis, Andrew. Introducing PROMIS – Next Generation HSFIS. www.cleverex.com/ShowArticle.asp?id=PROMIS. May 2001.

INFANT/TODDLER ENVIRONMENT RATING SCALE (ITERS), 1990

Authors:

Thelma Harms, Debby Cryer, Richard M. Clifford
Frank Porter
www.fpg.unc.edu

Publisher:

Teachers College Press
(800) 575-6566
www.teacherscollegepress.com

Initial Material Cost:

Assessment booklet: \$13
Score sheet: \$9
Video guide and training workbook: \$4
Video observations: \$59

Representativeness of Norming Sample: Research reported in manual does not include this information

Languages:

English¹, French, German, Italian

Type of Assessment:

Observation, with some direct caregiver assessment

Age Range and Administration Interval: For classrooms enrolling children from birth to 2_ years of age.

Personnel, Training, Administration, and Scoring Requirements:

Administered by a trained observer. The manual recommends that in group training, discussions of the ratings given by each observer should follow the trial observations. It also recommends that observers read the entire scale carefully, including the items and the Notes for Clarification, before attempting to rate a classroom. A block of at least 2 hours should be set aside for observation if an outside observer (anyone other than the caregiver) is doing the rating.

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (.65 or higher for internal consistency, test-retest and inter-rates reliability)

Validity: 3 (Validity is given in percentage, not correlation.)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The ITERS was adapted from the Early Childhood Environment Rating Scale (ECERS) and the Family Day Care Rating Scale (FDCRS) -- (see reviews in this section)—specifi-

cally for assessing the quality of out-of-home group care for infants and toddlers up to 30 months of age. The assessment includes 35 items divided into seven categories: Furnishings and

¹ The ITERS has also been translated into French, German, and Italian. Those interested may contact Thelma Harms or Debby Cryer.

Displays, Personal Care Routines, Listening and Talking, Learning Activities, Interactions, Program Structure, and Adult Needs. Each item is presented on a scale from 1 (Inadequate) to 7 (Excellent). The ITERS definition of the day care environment encompasses the organization of space, interaction, activities, schedule for children, and provisions for staff and parents. This definition of quality is consistent with the Criteria for Quality Early Childhood Programs stated by the National Academy of Early Childhood Programs and with the Child Development Associate (CDA) requirements.

Uses of Information: The ITERS provides information for development of the Individualized Family Service Plan and individual programming changes for specific children. It can be used by caregiving staff for self-assessment, by directors as a program-quality measure for planning program improvement, by agency staff for monitoring, and by parents concerned about quality care for their infants and toddlers. Key words in the scoring sheets can be used to make plans for improvement or compare scores.

Reliability: (1) Internal consistency (Cronbach's alpha) is .83, although subscale scores varied substantially. (2) Test-retest reliability (Spearman correlation), with a 3- to 4-week interval between tests, was .79 on the overall scale, with subscale scores ranging from .58 to .76. (3) Interrater reliability (Spearman correlation coefficient) was .84, with subscale scores ranging from .58 to .89.

Validity: (1) Concurrent validity: a test for concurrent validity found an 83 percent agreement between the categorizations of the quality of infant/toddler programs in 12 classrooms using ITERS and using expert evaluations.

Method of Scoring: The ITERS provides a score sheet that records individual item scores and a total score. The score sheet has space for the observer to briefly indicate the reasons for each of the scores by including key words describing what was observed. A sample completed score sheet is included in the manual. The Profile sheet permits a graphic representation of the scoring information. It can be used to compare areas of strengths and weaknesses, and to select items and subscales to target for improvement. The profiles for at least two observations can be plotted side by side to depict changes visually. Sample profile forms are available on the Frank Porter Graham Child Development Center website: www.fpg.unc.edu.

Interpretability: Full instructions for using the scale, plus notes clarifying selected scale items are included. However, the manual provides no information about interpreting the results of the observation.

Training Support: Video Observations and a Video Guide and Training Workbook are available for an additional cost. The Video Observations package demonstrates how to use the Infant-Toddler Environment Rating Scale. Each training package contains an interactive videotape and an Instructor's Guide, which explains how to present the various training activities and provides answers and explanations for any questions that

may arise. A 16-page Video Guide and Training Workbook, containing training activities, is sold separately. The website: www.fpg.unc.edu also includes helpful information.

Adaptations/Special Instructions for Individuals with Disabilities: The rating scale includes an item to rate accommodations for exceptional children.

Report Preparation Support: Not described in the manual

References:

Thelma Harms, Debby Cryer, and Richard M. Clifford. *Infant-Toddler Environment Rating Scale*. New York: Teachers College Press, 1990.

Thelma Harms, Debby Cryer, and Richard M.

Clifford. *Video Observations for the Infant-Toddler Environment Rating Scale*. New York: Teachers College Press, 1991.

Thelma Harms, Debby Cryer, and Richard M. Clifford. *Video Guide and Training Workbook for the Infant-Toddler Environment Rating Scale*. New York: Teachers College Press, 1991

Clifford, R. M., Russell, S., Fleming, J., Peisner, E. S., Harms, T., & Cryer, D. *Infant/Toddler Environment Rating Scale: Reliability and Validity Study-Final Report*. Chapel Hill, NC: Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill, 1989.

www.fpg.unc.edu

NATIONAL ASSOCIATION FOR THE EDUCATION OF YOUNG CHILDREN (NAEYC) ACCREDITATION CRITERIA, 1998

Authors:

National Association for the Education of Young Children (NAEYC)

Publisher:

NAEYC

(202) 232-8777 or (800) 424-2460

www.naeyc.org

Initial Material Cost:

Book: \$15

Accreditation fee varies by number of children:

60 or fewer children: \$575

61 to 120 children: \$900

121 to 240 children: \$1,100

241 to 360 children: \$1,250

\$200 for each additional 120 children

Representativeness of Norming Sample: Not applicable

Languages:

English

Type of Assessment:

Program staff self-report, reviewed by NAEYC validators if accreditation is pursued

Age Range and Administration Interval: Programs for young children of all ages. Accreditation good for 3 years.

Personnel, Training, Administration, and Scoring Requirements:

Not applicable

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 3 (95 percent agreement among reviewers)

Validity: Not applicable

Norming Sample Characteristics: Not applicable

Ease of Administration and Scoring: 1 (not described)

Description: The NAEYC accreditation system is designed to assist early childhood program personnel in making real and lasting improvements in quality of care and education, and to recognize programs that are in substantial compliance with the criteria for high-quality programs. NAEYC evaluates programs on the basis of their:

- Interaction among teachers and children
- Teaching
- Structure and processes used to facilitate the relationships among teachers and families
- Staff qualification and professional development opportunities
- Administration attention to the needs and desires of children, families, and staff
- Level of staffing

- Indoor and outdoor physical environment that facilitate learning

- Health and safety standards and practices

- Nutrition and food services

- Evaluation of program effectiveness.

NAEYC accreditation is a three-step process. The first step involves a self-evaluation, corrective actions, and completing the program description form that describes the program's compliance with the NAEYC criteria. The second step involves an on-site visit by NAEYC-trained validators to validate the accuracy of the information on the program description form. The third step involves the accreditation decision-making process by a three-person commission.

To be eligible for accreditation, the program must be licensed by the appropriate state/local agency or, if exempt from licensing, demonstrate compliance; serve at least 10 children younger than age 8; have at least two adults present at all times; and have been in operation for at least one year.

Uses of Information: Accredited programs may advertise their compliance with NAEYC standards. NAEYC provides accredited programs with a copy of the Commission Decision Report and their original validated program description, promotional materials, and a certificate indicating the expiration date of the accreditation. NAEYC provides programs with "deferred accreditation," along with the specific reason for the deferral and recommendations for improvement. Programs can use the information to make improvements

and request an additional on-site validation.

Reliability: NAEYC reports a 95 percent or higher agreement among commissioners reviewing the same program.

Validity: None described.

Method of Scoring: The three-person accreditation committee decides whether to grant accreditation or to defer accreditation until improvements can be made or additional information is obtained. The decision is based on the commissioners' professional judgment and not on a point system. The commissioners review the validated program decision, taking into account the context in which a program is operating and the overall impact of varying degrees of compliance for each component. To achieve accreditation, 100 percent compliance is not required; however, the program needs to be in substantial compliance with the accreditation criteria.

Interpretability: None described.

Training Support: Consultation by telephone and the Accreditation Criteria & Procedures of the National Academy of Early Childhood Programs and the Guide to Accreditation.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: Not applicable.

References:

Accreditation Criteria and Procedures of the National Association for the Education of Young Children, 1998 Edition. Washington, DC:

NAEYC.

Guide to Accreditation by the National Association for the Education of Young Children, 1998 Edition. Washington, DC: NAEYC.

Bredekamp, Sue and Barbara Willer (Eds.).

NAEYC Accreditation: A Decade of Learning and Years Ahead. Washington, DC: NAEYC, 1996.

Ethiel, Nancy (Ed.). Reflections on NAEYC Accreditation: Lessons Learned and Goals for the Future. Washington, DC: NAEYC, 1996.

PROGRAM IMPLEMENTATION CHECKLIST AND RATING SCALES DEVELOPED FOR THE NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT, 1997

Authors:

Ellen Kisker, Diane Paulsell, John Love, and Helen Raikes

Publisher:

Mathematica Policy Research, Inc.
Contact Publications, 609-275-2350,
jallen@mathematica-mpr.com, or visit the website,
www.mathematica-mpr.com

Initial Material Cost:

None

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Program staff self-report

Age Range and Administration Interval: Not applicable.

Personnel, Training, Administration, and Scoring Requirements:

Not applicable

Summary

Initial Material Cost: 1 (<\$100)

Reliability: 1 (none described)

Validity: Content validity established in relation to Head Start Program Performance Standards

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The program implementation checklist and rating scales developed for the national Early Head Start Research and Evaluation project were designed to guide the collection and organization of information related to 25 key elements of the Head Start Program Performance Standards and Early Head Start program guidelines. The checklist can be used to guide the collection of information related to the key program elements, and the rating scales can be used to assess how fully the program has implemented each key element, how fully key program areas are being implemented, and how fully the program is being implemented overall.

The checklist contains 39 general criteria with references to the performance standards, as well as specific indicators for each general criterion. Completing the checklist requires collecting information from staff, parents, and program records. There are five ratings scales, one each for early childhood development and health services, family partnerships, staff development, community partnerships, and management systems. Each rating scale has multiple dimensions with ratings from 1 to 5. A rating of 4 indicates full implementation and a rating of 5 indicates enhanced implementation.

Uses of Information: The checklist and rat-

ing scales are intended to help Head Start programs serving pregnant women and families with infants and toddlers identify areas that need improvements in order to continue to provide high-quality comprehensive services that meet Head Start performance standards and other regulations. They may be useful for organizing information in preparation for Head Start Bureau monitoring visits.

Reliability: Reliability has not been established in the usual sense. However, an informal assessment conducted by Head Start Bureau monitoring staff concluded that the assessments of “full” or “enhanced” implementation on these rating scales were consistent with results of in-depth monitoring conducted by the Bureau.

Validity: Content validity was established by reviewing the specific criteria for determining the rating on each dimension with representatives of the Head Start Bureau and the Early Head Start National Resource Center. A form of predictive validity was assessed in the national Early Head Start Research and Evaluation project, which found that programs that were rated as fully implemented achieved a stronger pattern of impacts on children and families.

Method of Scoring: The Early Head Start evaluation team developed overall ratings of each area and for the program overall by having multiple team members rate each program independently, meet to discuss any discrepancies in ratings,

and agree on a consensus rating for each dimension. The team also created summary ratings of each area and for the program overall. To be rated fully implemented overall, a program had to receive a rating of 4 or 5 on most dimensions rated. This process could be followed by program staff who wanted to develop summary ratings.

Interpretability: The results obtained from the ratings of program implementation are readily interpretable by programs serving families with infants and toddlers to show areas of programmatic strengths and weaknesses. Because the scales are tied to key dimensions of the performance standards, program management and staff can see ways to focus program improvement efforts.

Training Support: The Early Head Start evaluation’s final implementation report, *Pathways to Quality* (ACF 2002) describes the use of the checklist and rating scales in the national Early Head Start Research and Evaluation project. Copies of the rating are available in the report.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: Not applicable.

References:

Administration for Children and Families. *Pathways to Quality and Full Implementation in Early Head Start Programs*. Washington, DC: U.S. Department of Health and Human Services, 2003.

PROGRAM REVIEW INSTRUMENT FOR SYSTEMS MONITORING (PRISM), 2002

Authors:

American Institute for Research

Publisher:

Head Start Bureau

(866) 763-6481

puborder@headstartinfo.org

Initial Material Cost:

No costs

Representativeness of Norming Sample: Not applicable

Languages:

English

Type of Assessment:

Comprehensive qualitative and quantitative assessment of program-level activities

Age Range and Administration Interval: For programs serving families with children birth to age 5.

Personnel, Training, Administration, and Scoring Requirements:

Administered by Head Start Bureau personnel for program monitoring; can be used by program staff for self-review

Summary

Initial Material Cost: 1 (<\$100) Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 1 (none described)

Description: The Program Review Instrument for Systems Monitoring (PRISM) is both the instrument and the process used by the Administration on Children, Youth and Families to monitor Head Start programs to ensure compliance with program performance standards and other applicable regulations. The PRISM instruments are based on 17 Core Questions—9 that focus on program services and 8 that focus on management systems. The PRISM review team uses a variety of methods to learn about a program. The team members interview staff, parents, community partners, and Policy Council and governing body members, individually and in

groups. They observe classrooms and family child care settings, and conduct home visits. They also complete fiscal, health and safety, and bus ride checklists. PRISM review decisions are done through consensus about the program quality.

Uses of Information: The assessment is intended to help Head Start programs identify areas that need improvements in order to continue to provide high-quality comprehensive services that meet Head Start performance standards and other regulations. The exact remedies are left to the program.

Reliability: Not described.

Validity: Not described.

Method of Scoring: Throughout the review visit, the PRISM review team, guided by the federal team leader, holds formal and informal briefings with grantee staff to report on information team members witnessed, heard, and read. During these briefings, the grantee staff members are able to provide input on the findings. At the end of the review, the review team holds an exit meeting and summarizes its findings in three areas—Child Development and Health Services, Family and Community Partnerships, and Program Design and Management. The review team will meet to share and analyze information collected during the visit. The team will work toward building consensus on issues related to the Core Questions. The review team members will then draft a three-part report—one for each area—that summarizes the program’s strengths and areas of concerns, reviews decisions by Core Questions, and lists findings requiring corrective action. The team leader, after consulting with other Regional Office staff, will use the draft

report to prepare the Official Report.

Interpretability: No information.

Training Support: The Head Start publication, Partnership for Quality: A Grantee Guide to PRISM 2002 discusses what PRISM is and the PRISM review process, and provides suggestions on what programs can do to prepare for the review.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable. The PRISM review criteria include evaluation of areas related to services for children with disabilities.

Report Preparation Support: Not applicable.

References:

PRISM: Program Review Instrument for Systems Monitoring of Head Start and Early Head Start Grantees—All Instruments.

PRISM: Program Review Instrument for Systems Monitoring of Head Start and Early Head Start Grantees—Partnerships for Quality: A Grantee Guide to PRISM 2002.